Université de Montréal

Développements et applications de méthodes computationnelles pour l'étude de l'agrégation des protéines amyloïdes

par
Sébastien Côté

Département de physique
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)
en physique

Août, 2015

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée:

Développements et applications de méthodes computationnelles pour l'étude de l'agrégation des protéines amyloïdes

présentée par:

Sébastien Côté

a été évaluée par un jury composé des personnes suivantes:

Jean-Yves Lapointe,	président-rapporteur
Normand Mousseau,	directeur de recherche
Rémy Sauvé,	membre du jury
Régis Pomès,	examinateur externe
Karen Waldron,	représentant du doyen de la FES

Thèse acceptée le: . . . . . .

RÉSUMÉ

Les protéines sont au coeur de la vie. Ce sont d'incroyables nanomachines moléculaires spécialisées et améliorées par des millions d'années d'évolution pour des fonctions bien définies dans la cellule. La structure des protéines, c'est-à-dire l'arrangement tridimensionnel de leurs atomes, est intimement liée à leurs fonctions. L'absence apparente de structure pour certaines protéines est aussi de plus en plus reconnue comme étant tout aussi cruciale. Les protéines amyloïdes en sont un exemple marquant : elles adoptent un ensemble de structures variées difficilement observables expérimentalement qui sont associées à des maladies neurodégénératives.

Cette thèse, dans un premier temps, porte sur l'étude structurelle des protéines amyloïdes bêta-amyloïde (Alzheimer) et huntingtine (Huntington) lors de leur processus de repliement et d'auto-assemblage. Les résultats obtenus permettent de décrire avec une résolution atomique les interactions des ensembles structurels de ces deux protéines.

Concernant la protéine bêta-amyloïde (Aβ), nos résultats identifient des différences structurelles significatives entre trois de ses formes physiologiques durant ses premières étapes d'auto-assemblage en environnement aqueux. Nous avons ensuite comparé ces résultats avec ceux obtenus au cours des dernières années par d'autres groupes de recherche avec des protocoles expérimentaux et de simulations variés. Des tendances claires émergent de notre comparaison quant à l'influence de la forme physiologique de Aβ sur son ensemble structurel durant ses premières étapes d'auto-assemblage. L'identification des propriétés structurelles différentes rationalise l'origine de leurs propriétés d'agrégation distinctes. Par ailleurs, l'identification des propriétés structurelles communes offrent des cibles potentielles pour des agents thérapeutiques empêchant la formation des oligomères responsables de la neurotoxicité.

Concernant la protéine huntingtine, nous avons élucidé l'ensemble structurel de sa région fonctionnelle située à son N-terminal en environnement aqueux et membranaire. En accord avec les données expérimentales disponibles, nos résultats sur son repliement en environnement aqueux révèlent les interactions dominantes ainsi que l'influence sur celles-ci des régions adjacentes à la région fonctionnelle. Nous avons aussi caractérisé la stabilité et la croissance de structures nanotubulaires qui sont des candidats potentiels aux chemins d'auto-assemblage de la région amyloïde de huntingtine. Par ailleurs, nous avons également élaboré, avec un groupe d'expérimentateurs, un modèle détaillé illustrant les principales interactions responsables du rôle

d'ancre membranaire de la région N-terminal, qui sert à contrôler la localisation de huntingtine dans la cellule.

Dans un deuxième temps, cette thèse porte sur le raffinement d'un modèle gros-grain (sOPEP) et sur le développement d'un nouveau modèle tout-atome (aaOPEP) qui sont tous deux basés sur le champ de force gros-grain OPEP, couramment utilisé pour l'étude du repliement des protéines et de l'agrégation des protéines amyloïdes. L'optimisation de ces modèles a été effectuée dans le but d'améliorer les prédictions *de novo* de la structure de peptides par la méthode PEP-FOLD. Par ailleurs, les modèles OPEP, sOPEP et aaOPEP ont été inclus dans un nouveau code de dynamique moléculaire très flexible afin de grandement simplifier leurs développements futurs.

Mots clés: protéine, amyloïde, bêta-amyloïde, huntingtine, interactions protéine–membrane, aaOPEP, opep_sim.

ABSTRACT

Proteins are at the center of life. They are formidable molecular nanomachines specialized and optimized during million years of evolution for well-defined functions in the cell. The structure of proteins, meaning the tridimensional setting of their atoms, is closely related to their function. Absence of structure for a subset of proteins is also recognized to be as crucial. Amyloid proteins is a striking example : they fold into an ensemble of various structures hardly observable experimentally that are associated with neurodegenerative diseases.

This thesis, firstly, is on the study of the structural ensemble of the amyloid proteins amyloid-beta (Alzheimer) and huntingtin (Huntington) during their folding and aggregation. Our results describe in details, with an atomic resolution, the characteristic interactions present in the structural ensemble of these two proteins.

Concerning the amyloid-beta protein (Aβ), our results show the structural differences between three of its physiological forms during its first aggregation steps in an aqueous environment. We have then compared these results with those obtained during the past few years by several other research groups using various experimental and simulation protocols. Clear trends come out of this comparison regarding the influence of Aβ physiological form on its structural ensemble during its first aggregation steps. Their distinct aggregation pathways are rationalized by the identified differences. For their part, the identified similarities offer targets for therapeutical compounds disrupting the aggregation of the neurotoxic oligomers.

Concerning the huntingtin protein, we identify the structural ensemble of its functional region at its N-terminal in an aqueous environment and in a phospholipid membrane. In agreement with the available experimental results on the global structure of this region in aqueous solution, our results reveal the dominant interactions, at an atomic precision, in its structural ensemble as well as the influence of its neighboring regions. We have also characterized the stability and the growth of nanotube-like structures that could occur during the aggregation of the amyloid region of huntingtin. Moreover, we have developed, in collaboration with a group of experimentalists, a precise model describing the main membrane interactions of huntingtin N-terminal, which serves as a membrane anchor that controls the localization of huntingtin in the cell.

Secondly, this thesis is on the refinement of a coarse-grained model (sOPEP) and on the development of a new all-atom model (aaOPEP) that are both based on the coarse-grained OPEP

force field, commonly used to study protein folding and amyloid protein aggregation. The goal behind the optimization of these models is to improve the *de novo* structure prediction of the PEP-FOLD method. These three models – OPEP, sOPEP and aaOPEP – are now also implemented in a new molecular dynamics software that we have developed specifically to greatly ease their future developments.

Keywords: protein, amyloid, amyloid-beta, huntingtin, protein–membrane interactions, aaOPEP, opep_sim.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

LISTE DES ANNEXES

LISTE DES SIGLES

Aβ	Protéine bêta-amyloïde
DC	Spectroscopie par dichroïsme circulaire
DM	Dynamique moléculaire
DRX	Diffractométrie de rayons X
pont-H	Pont hydrogène
HREX	Dynamique moléculaire d'échange de repliques Hamiltonien
HT-REMD	Dynamique moléculaire d'échanges de répliques en température et Hamiltonien
Htt	Protéine huntingtine
Htt^{NT}	Amino-terminal de la protéine Huntingtin
IRTF	Spectroscopie infrarouge à transformée de Fourier
MétaD	Métadynamique
MQ	Mécanique quantique
MM	Mécanique moléculaire
MTS	Multiple timestep
NPT	Ensemble avec nombre de particules, pression et température constants
NVE	Ensemble microcanonique
NVT	Ensemble canonique
REMD	Dynamique moléculaire d'échanges de répliques
RMN	Résonance magnétique nucléaire
T-REMD	Dynamique moléculaire d'échanges de répliques en température
vdw	Interactions de van der Waals

(dédicace) À ma famille qui a toujours
soutenu ma curiosité scientifique.
Et, SLOLOLZ B & m^2.

REMERCIEMENTS

Je remercie chaleureusement ma famille, ma conjointe et mes amis d'avoir toujours supporté et entretenu ma curiosité scientifique. Sans eux, je ne me serais pas rendu jusqu'ici.

Je remercie grandement le professeur Normand Mousseau de m'avoir permis d'intégrer son groupe de recherche. Je dois à Normand une part importante de mon enrichissement scientifique, professionnel et personnel durant les six dernières années. J'ai beaucoup appris de lui. Il m'a donné l'opportunité de développer idées, projets et collaborations, lesquels ont permis plusieurs stages de recherches et conférences. Sous sa supervision, j'ai pu repousser mes limites.

Je remercie également les professeurs Pierre Tufféry de l'INSERM à Paris et Guanghong Wei de l'Université Fudan à Shanghai pour m'avoir accueilli dans leur laboratoire respectif. Ces stages ont été très formateurs pour moi : mes travaux sur aaOPEP et sur huntingtine découle ultimement de mon stage chez Pierre et Guanghong, respectivement. J'ai aimé collaborer directement avec vous et j'espère que nous nous reverrons.

Je souligne aussi l'apport essentiel de nos autres collaborateurs à mes activités de recherche. Le professeur Phillipe Derreumaux de l'Université Paris Diderot a été l'instigateur de mes travaux sur la protéine Bêta-amyloïde et il a permis d'élever la qualité de mes premiers pas en recherche. De plus, j'ai particulièrement apprécié travailler avec le professeur Burkhard Bechinger et Evgeniy Salnikov de l'Université Strasbourg sur les interactions huntingtine–membrane. Nous avons élaboré ensemble un modèle décrivant plus précisément ces interactions en combinant résultats expérimentaux et simulations.

Je suis aussi reconnaissant envers les divers organismes qui m'ont témoigné leur confiance en finançant mes activités de recherche : le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), le Fonds de recherche du Québec – Nature et technologies (FRQNT), le Groupe d'étude des protéines membranaires (GÉPROM) et la Faculté des études supérieures et postdoctorales de l'Université de Montréal (FESP).

J'ai finalement eu l'honneur de côtoyer des collègues sensationnels tout au long de ces six années : Laurent Karim Béland, Vincent Binette, Saïd Bouzakraoui, Peter Brommer, Guo Cong, Lilianne Dupuis, Cindie Eugène, Jean-François Joly, Laurent Jutras-Dubé, Gawonou Kokou N'Tsouaglo, Céline Labbé, Rozita Laghaei, Gautier Moroy, Jessica Nasica-Labouze, Géraldine Picard, Qi Ruxi, Cao Siqin, Jean-François St-Pierre, Mickaël Trochet et Su Wei. Je remercie

spécialement Lillianne, Rozita et Jessica pour leur accueil et leur support lors de mon arrivée dans le groupe Biophysique, Vincent pour son excellence en recherche lors de notre collaboration sur huntingtine, Géraldine pour sa joie de vivre, ainsi que Guo Cong, Cao Siqin et Su Wei pour la découverte de la Chine et de ses délices culinaires.

CHAPITRE 1

INTRODUCTION

« Human : the final frontier »[1], nous sommes facilement fascinés par l'immensité du cosmos tout en oubliant la complexité de l'humain à l'échelle cellulaire. Le corps humain possède en effet plus de cellules qu'il n'y a de galaxies dans l'Univers. À lui seul, le nombre de protéines – l'unité fonctionnelle des cellules – rivalise avec le nombre d'étoiles dans l'Univers. Ce qui est d'autant plus fascinant, c'est que toutes ces cellules et toutes ces protéines agissent de concert grâce à plus de trois milliards d'années d'évolution pour nous donner la vie. Les biophysiciens s'attèlent donc à quantifier le fonctionnement des cellules en faisant le pont entre la physique, la biologie et la chimie, ainsi qu'entre l'expérience, la théorie et la modélisation numérique [65, 139, 262, 411].

Au centre des fonctions cellulaires, les protéines jouent le rôle fondamental de nanomachines moléculaires. Les enzymes, par exemple, catalysent des réactions chimiques qui permettent entre autres la copie de notre ADN lors de la division cellulaire (ADN polymérases), la contraction de nos muscles dont ceux de notre coeur (myosines), le bris des molécules lors de la digestion des aliments (peptidases) et la conversion de l'amidon en sucre[2] (amylases). Diverses enzymes sont aussi impliquées dans la respiration cellulaire qui permet, en quelque sorte, à nos cellules de convertir les aliments que l'on mange en énergie. D'autres protéines régulent l'activité d'un ensemble de protéines : la calmoduline, par exemple, active la contraction musculaire et l'activité d'une panoplie de protéines membranaires en présence de calcium. Les nombreuses protéines membranaires, quant à elles, permettent à la cellule de communiquer et d'échanger des composés essentiels dans le but de maintenir un équilibre avec son environnement. Par exemple, les récepteurs couplés à la protéine G sont essentiels pour transmettre un stimulus environnemental à l'intérieur de la cellule, ce qui en fait une cible pharmacologique particulièrement intéressante. La photosynthèse se produisant dans les plantes est aussi assurée par un complexe de protéines membranaires.

La structure des protéines, c'est-à-dire l'arrangement tridimensionnel de leurs atomes, est au coeur de leurs fonctions. Ce principe a pris forme depuis la détermination de la première structure tridimensionnelle de l'ADN en 1953 par James D. Watson, Francis Crick et Rosalind

[1] Adaption libre de « Space : the final frontier », Star Trek.

[2] Phénomène essentiel autant dans notre corps que lors de la fermentation de la bière !

Franklin [493][3] ainsi que de deux protéines – la myoglobine et l'hémoglobine – en 1957–1959 par John Kendrew et Max Perutz [60, 234] grâce à la diffractométrie de rayons X (DRX).[4] Les décennies suivantes ont vu le perfectionnement de la DRX et la naissance d'une nouvelle technique – la résonance magnétique nucléaire en solution (RMN) [503] – permettant la découverte de dizaines de milliers de structures tridimensionnelles de protéines supplémentaires. À ce jour, la plus importante banque de données de structures contient plus de 100 000 modèles atomistiques d'environ 20 000 protéines différentes [43].[5] Bien que ce chiffre semble impressionnant, le nombre total de protéines codées par le génome eucaryote est estimé à plusieurs millions [250] dont 20 000 à 25 000 pour le génome humain [196]. C'est ici que des approches de bio-informatique et biophysique computationnelle capables de prédire très rapidement la structure des protéines prennent tout leur sens. Dans cette thèse, je présente nos améliorations à la méthode PEP-FOLD [430] qui prédit, en quelques secondes, la structure de petites protéines seulement à partir de leur séquence d'acides aminés (Chapitre 13). Plus précisément, nous avons développé un protocole d'optimisation permettant l'amélioration de la fonction de pointage sOPEP [310], au coeur de cette méthode, qui évalue la qualité des modèles structurels produits. De plus, nous avons développé un nouveau modèle d'interaction physique (aaOPEP) [107] et un nouveau code de dynamique moléculaire (opep_sim) [105] permettant de raffiner la qualité structurelle des modèles produits par PEP-FOLD.

L'absence de structure est de plus en plus reconnue comme étant tout aussi cruciale [353, 477, 508]. Ces protéines intrinsèquement désordonnées en isolation transitent à des états ordonnés lors de situations spécifiques telles que la fixation sur une autre protéine ou sur une membrane lipidique, la présence de certaines conditions environnantes (concentrations ioniques et pH) et l'auto-assemblage en homo-oligomères. Les protéines amyloïdes, par exemple, sont principalement désordonnées en isolation, mais s'auto-assemblent en structures très ordonnées appelées fibres amyloïdes [339, 469, 501]. Elles peuvent aussi s'ordonner au contact d'une membrane lipidique [263, 264, 525]. Bien que désordonnés, les agrégats amorphes, quant à eux, ont un rôle fonctionnel dont l'origine reste inconnue dans plusieurs maladies neurodégénératives telles que l'Alzheimer (protéine bêta-amyloïde) [179, 417, 489] et Huntington (protéine huntingtine) [26, 540]. Or, la structure désordonnée de ces agrégats ainsi que leur variabilité struc-

[3] J. D. Watson et F. Crick ont partagé le prix Nobel de physiologie ou médecine en 1962 pour leur découverte.

[4] J. Kendrew et M. Perutz ont partagé le prix Nobel de chimie en 1962 pour leurs nombreux travaux sur les protéines globulaires.

[5] Nombre de protéines ayant une homologie de séquence en deçà de 30%.

turelle les rendent difficilement observables expérimentalement au niveau de l'atome par DRX et RMN. C'est ici que des approches de biophysique computationnelle capables d'échantillonner convenablement l'ensemble structurel de ces protéines permettent de compléter les observations expérimentales. Dans cette thèse, je présente mes recherches sur la caractérisation de l'ensemble structurel des protéines amyloïdes bêta-amyloïde (Chapitre 4) [108, 109] et huntingtine (Chapitre 8) [52, 106, 110, 111] en solution aqueuse et en environnement membranaire grâce à des méthodes *in silico*, par ordinateur, variées.

Dans les chapitres suivants, je résumerai tout d'abord la biochimie et la physique des protéines (Chapitre 2) ainsi que les diverses méthodologies *in silico* que j'ai utilisées (Chapitre 3). Ensuite, je mettrai en contexte mes travaux sur la protéine bêta-amyloïde (Chapitre 4) au niveau du monomère (Chapitre 5) et du dimère (Chapitre 6) ainsi qu'une synthèse récente des résultats expérimentaux et de simulations sur les petits agrégats de cette protéine (Chapitre 7). Puis, je mettrai en contexte mes travaux sur la protéine huntingtine (Chapitre 8) au niveau du monomère (Chapitre 9), des oligomères (Chapitre 10) et de ses interactions avec une membrane de phospholipides (Chapitres 11 et 12). Finalement, je présenterai le nouveau champ de force *all-atom* OPEP que j'ai conçu et paramétré pour le raffinement des prédictions structurelles de PEP-FOLD (Chapitre 13).

CHAPITRE 2

LA BIOCHIMIE ET LA PHYSIQUE DES PROTÉINES

Les protéines sont des molécules dont l'arrangement tridimensionnel des atomes leur confère des fonctions biologiques. Elles sont des nanomachines cruciales au fonctionnement des cellules et elles sont synthétisées par ces dernières. Dans certains cas, comme pour les protéines amyloïdes, elles transitent à un état ordonné seulement sous certaines conditions précises.

Dans ce chapitre, les aspects suivants seront présentés : la régulation de l'expression des protéines, leur structure atomique et les méthodes expérimentales permettant de la caractériser, les interactions physiques responsables de leur structure, les modèles de repliement des protéines ordonnées, et les caractéristiques des protéines désordonnées comme les protéines amyloïdes.

2.1 La vie des protéines

La synthèse et la régulation des protéines font intervenir plusieurs composantes de la cellule, qui est l'unité de base de tout organisme pluricellulaire comme les animaux et végétaux [191].

2.1.1 La cellule

Les organismes des règnes animal et végétal sont tous composés de cellules eucaryotes ayant une morphologie très similaire, dont la présence d'un noyau entouré par la membrane nucléaire (Figure 2.1). La cellule eucaryote est entourée par la membrane plasmique, laquelle est composée de phospholipides, qui sert de barrière entre la cellule et son milieu environnant. Dans cette membrane se trouve un nombre important de protéines qui permettent à la cellule d'échanger des composés et de communiquer de façon très régulée avec l'extérieur. À l'intérieur d'une cellule animale typique, on retrouve des sous-structures entourées d'une membrane – les organites – qui baignent dans le cytosol : les mitochondries sont en quelque sorte des usines à énergie, le réticulum endoplasmique exporte les protéines dans des vésicules vers l'extérieur de la cellule ou dans une membrane pour les protéines membranaires, l'appareil de Golgi peut induire des modifications supplémentaires à la structure moléculaire des protéines, les lysosomes sont des vésicules ayant un intérieur très acide contenant des enzymes qui permettent la dégradation de protéines, et le peroxysome permettent la détoxification de la cellule en dégradant le peroxyde

d'hydrogène produit par les mitochondries. L'intégrité et la structure d'une cellule sont maintenues en place par le cytosquelette, lequel est aussi important dans le transport intracellulaire des protéines.

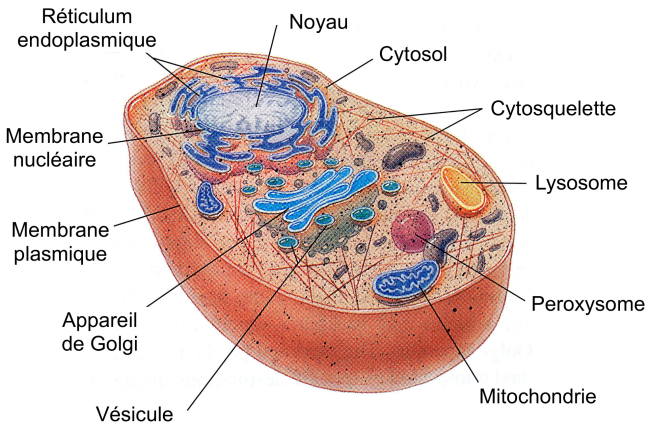


Figure 2.1 – Schématisation d'une cellule eucaryote typique du règne animal. Les organites principaux – les mitochondries, le réticulum endoplasmique, l'appareil de Golgi, les lysosomes et les peroxysomes – sont illustrés. Image adaptée [191].

2.1.2 La synthèse des protéines

Tout débute dans le noyau qui contient l'ensemble des gènes – le génome – lequel indique, entre autres, la procédure pour synthétiser l'ensemble des protéines – le protéome – présent dans notre organisme (Figure 2.1). Le génome est donc une bibliothèque de 0.8 gigabytes[1] avec tous les modes d'emploi pour créer le protéome. L'information génétique est contenue dans des molécules d'acide désoxyribonucléique (ADN) formées de nucléotides. Des segments d'ADN – les gènes – peuvent être transcrits en acide ribonucléique messager (ARNm) par certaines protéines spécialisées. L'ARNm diffuse en dehors du noyau pour rejoindre un des nombreux ribosomes qui décode ensuite l'ARNm et procède à la synthèse de la protéine associée. Une fois synthétisée, la protéine peut subir dans le Golgi des modifications dites post-traductionnelles qui permettent de contrôler sa localisation et ses fonctions dans la cellule.

[1] Espace disque nécessaire pour contenir l'information du génome d'un humain (1 byte pour 4 paires de bases et notre génome en a environ 3,2 milliards).

2.2 La structure atomique des protéines

Durant et après sa synthèse, chaque protéine se replie en une structure tridimensionnelle spécifique qui est intimement reliée à ses fonctions. La structure des protéines est décrite par quatre niveaux d'organisation : primaire, secondaire, tertiaire et quaternaire (Figure 2.2) [64].

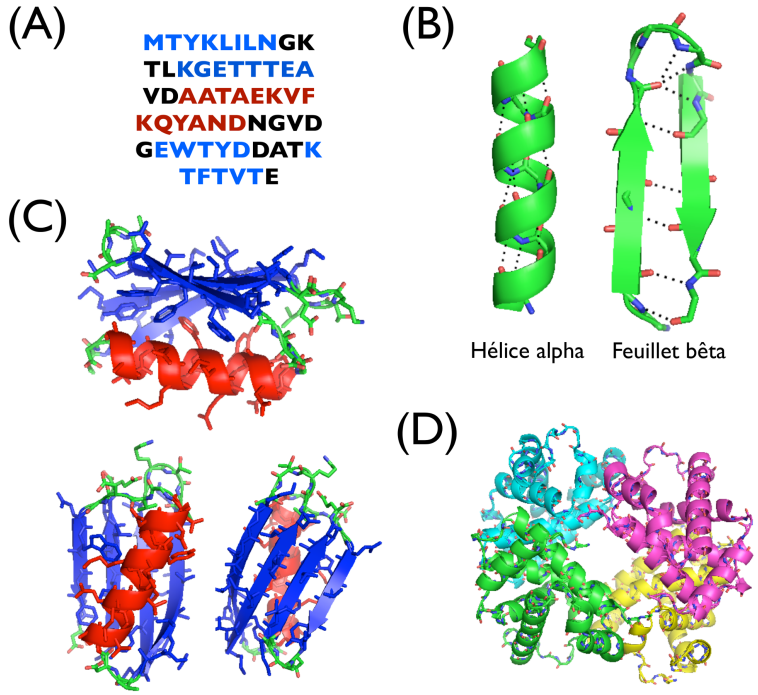


Figure 2.2 – Les quatre niveaux d'organisation de la structure des protéines : structures primaire, secondaire, tertiaire et quaternaire. En exemple, le domaine B1 liant l'immunoglobuline de la protéine G du streptocoque (PDB : 1PGB) avec ses niveaux (A) primaire, (B) secondaire, et (C) tertiaire. (D) La structure quaternaire de l'hémoglobine (PDB : 4HHB) où chaque chaîne polypeptidique est colorée différemment.

2.2.1 La structure primaire

La structure primaire est la séquence linéaire d'acides aminés qui composent la chaîne polypeptidique d'une protéine (Figure 2.2A). Le ribosome les assemble par une réaction de condensation, soit la formation du lien peptidique entre l'acide carboxylique (COO^-) et l'amine (NH_3^+) de deux acides aminés consécutifs. Par conséquent, le premier acide aminé d'une chaîne polypeptidique débute par un groupement amine et porte ainsi le nom de N-terminal, tandis que le dernier acide aminé se termine par un acide carboxylique et porte ainsi le nom de C-terminal.

Une protéine peut contenir de quelques dizaines à plus de deux mille acides aminés.

Dans chaque acide aminé, certains atomes – N, H_N, C_α, H_α, C et O – participent au squelette de la protéine, tandis que les autres forment ce qui est appelé la chaîne latérale (Figure 2.3). Il existe seulement 20 types d'acides aminés standards, c'est-à-dire sans modification post-traductionnelle, qui possèdent des propriétés physico-chimiques différentes, mais qui sont tous composés des mêmes cinq atomes (carbone, azote, oxygène, hydrogène et soufre).[2] Ils sont séparés en quatre grandes classes selon les propriétés de leur chaîne latérale : polaire (Asn, Cys, Gln, Ser, Thr, Trp et Tyr), non-polaire (Ala, Phe, Pro, Ile, Leu, Met et Val), chargé[3] positivement (Arg, His et Lys) ou négativement (Asp et Glu) et spécial (Gly). Il existe également une classification basée sur la structure atomique des acides aminés : aliphatique (Ala, Gly, Ile, Leu, Pro, Val), aromatique (Phe, Trp et Tyr), contenant un soufre (Cys et Met), alcool (Ser et Thr), base (Arg, His et Lys) et amide (Asn et Gln).

Les autres niveaux de la structure sont une conséquence directe des propriétés physico-chimiques de la séquence d'acides aminés particulière à chaque protéine. La structure globale et les fonctions d'une protéine sont donc intrinsèquement codées dans sa structure primaire. La découverte de l'ensemble des règles régissant la relation séquence–structure–fonction reste le Saint-Graal dans notre compréhension des protéines.

2.2.2 La structure secondaire

La structure secondaire est la structure locale du squelette de la protéine et il en existe trois grandes classes : les feuillets, les hélices et les coudes (Figure 2.2B).

La structure locale du squelette de la protéine entre deux acides aminés adjacents est décrite par trois angles dièdres : ϕ (autour de C–N–C_α–C), ψ (autour de N–C_α–C–N) et ω (autour de C_α–C–N–C_α) tel qu'illustré à la Figure 2.4. Les atomes autour de la liaison peptidique (C^i–N^{i+1}) entre deux acides aminés consécutifs – C_α^i, C^i, O^i, N^{i+1}, H_N^{i+1} et C_α^{i+1} – forment un plan appelé le plan peptidique. L'angle dièdre ω de ce plan est assez fixe autour de deux valeurs : 0° (*cis*) ou 180° (*trans*) où la configuration *trans* est la plus répandue *in vivo* parce qu'énergéti-

[2] Bien que vingt types peut sembler bien peu, ceci permet néanmoins 20^N combinaisons possibles pour une séquence de N acides aminés. Il s'agit, en effet, d'un nombre astronomique : $20^{20} \approx 1 \cdot 10^{26}$ combinaisons possibles pour une très petite protéine de seulement 20 acides aminés. C'est un chiffre plus imposant que celui de l'âge de l'Univers en seconde : $4 \cdot 10^{17}$ s.

[3] Pour un pH neutre de 7. De plus, notons que le ratio d'histidine ayant une charge totale positive versus neutre est de 1 pour 10 pour un pH neutre.

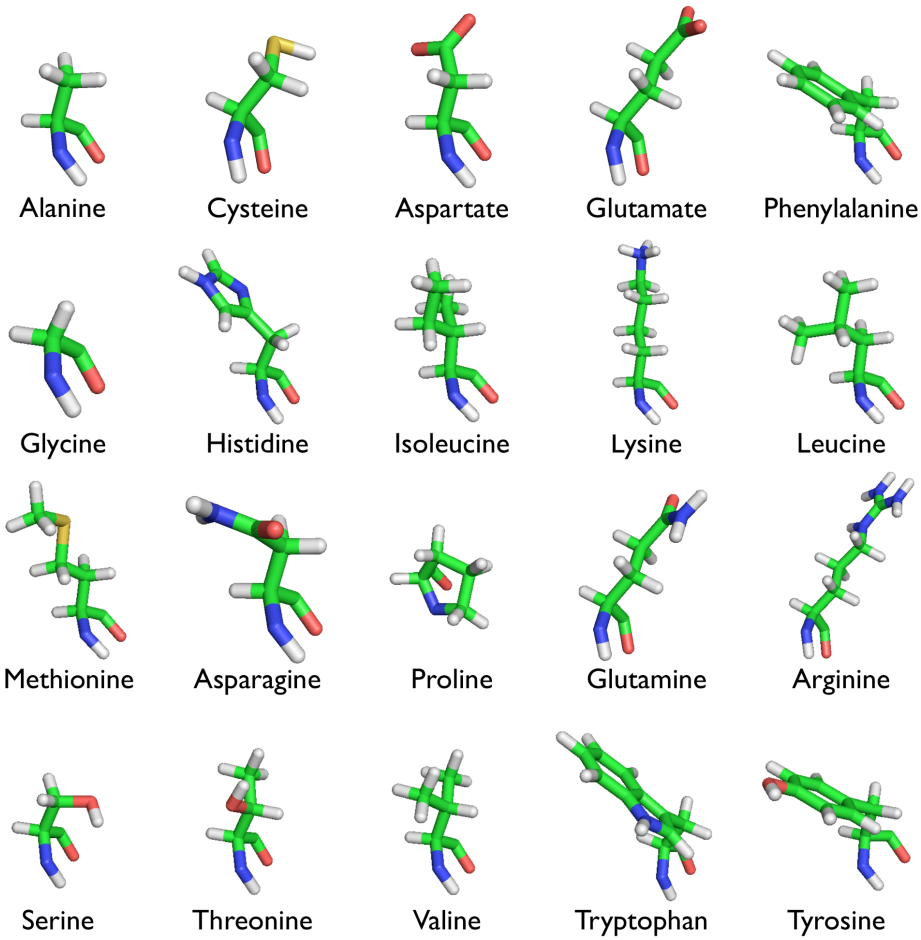


Figure 2.3 – Les 20 acides aminés standards. Les atomes d'hydrogène, d'azote, de carbone, d'oxygène et de soufre sont respectivement colorés en blanc, bleu, vert, rouge et jaune. Les images ont été produites avec le système de graphique moléculaire PyMOL.

quement plus favorable. Les angles dièdres ϕ et ψ entre deux plans peptidiques consécutifs sont restreints à des régions spécifiques tel qu'illustré par le diagramme de Ramachandran [389]. Ceci est dû aux contraintes stériques imposées par les atomes des chaînes latérales consécutives.

Trois régions principales sont identifiées sur le diagramme de Ramachandran : étendue (brins-β), hélicoïdale droite (hélices-α, -3_{10} et -π) et hélicoïdale gauche (hélices-α_L). Les valeurs (ϕ,ψ) des coudes sont principalement concentrées entre les régions étendue et hélicoïdale droite.

Au moins deux brins-β doivent s'associer en une structure tertiaire afin de se stabiliser en formant un feuillet-β parallèle ou anti-parallèle selon l'orientation relative des deux brins-β (Figure 2.5). L'arrangement anti-parallèle est le plus stable énergétiquement, car la direction des

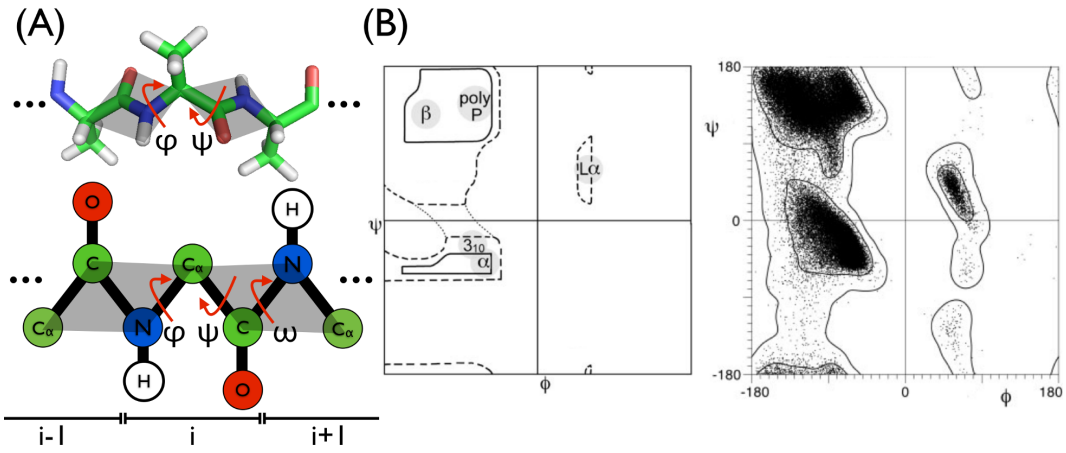


Figure 2.4 – (A) Définition graphique des angles dièdres ϕ et ψ et ω du squelette de la protéine. Les atomes d'hydrogène, d'azote, de carbone et d'oxygène sont respectivement colorés en blanc, bleu, vert et rouge. Le segment tri-alanine a été produit avec le système de graphique moléculaire PyMOL. (B) Diagramme de Ramachandran montrant les valeurs d'angles dièdres ϕ et ψ accessibles ainsi que leur fréquence d'apparition dans les structures des protéines de la Protéine Data Bank, images adaptées [289].

ponts-H est optimale. Ceci ne peut pas se produire dans l'arrangement parallèle : la direction des ponts-H doit être déviée afin d'éviter des collisions stériques entre les chaînes latérales des deux brins-β. La longueur des brins-β est d'environ 6 acides aminés consécutifs. Les feuillets-β consistent en 2 à 15 brins-β et sont tordus vers la droite.

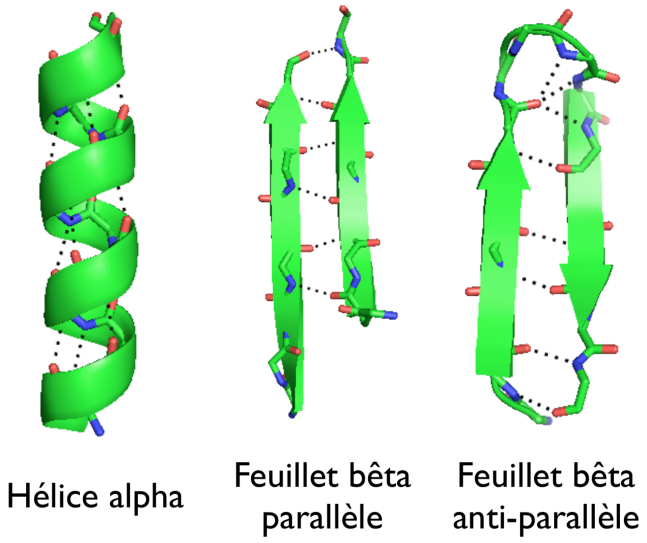


Figure 2.5 – La topologie des ponts-H stabilisants les feuillets-β et les hélices-α. Les ponts-H sont illustrés par les traits noirs pointillés.

Les hélices droites se présentent en trois formes qui sont caractérisées par une topologie différente de leurs ponts-H et par un nombre différent d'acides aminés par tour : hélices-α, hélices-3_{10} et hélices-π telles qu'illustrées à la Figure 2.5.[4] Les hélices-α possèdent 3,6 acides aminés par tour d'hélice et un pont-H entre chaque paire d'acides aminés espacée par 4 acides aminés (i/i+4) dans la structure primaire. Elles sont les plus communes et les plus longues, ayant de 4 à 40 acides aminés consécutifs. Les hélices-3_{10} et -π possèdent respectivement 3 et 4,1 acides aminés par tour d'hélice. Leurs ponts-H se forment respectivement entre chaque paire d'acides aminés espacés par 3 (i/i+3) et 5 (i/i+5) acides aminés. Ces hélices sont plus rares et souvent très courtes ayant de 3 à 7 acides aminés consécutifs. Elles peuvent aussi être présentes dans une hélice-α afin de permettre l'accommodation d'un acide aminé supplémentaire ou manquant. Les hélices gauches, quant à elles, sont beaucoup plus rares et se retrouvent principalement dans les coudes ou dans les hélices très courtes. Une caractéristique commune des hélices est qu'elles forment un dipôle dû à l'orientation de leurs ponts-H.[5]

La proportion des types d'acides aminés est différente autant dans les hélices que dans les feuillets. Deux acides aminés sont particulièrement spéciaux en ce qui concerne la structure secondaire : la glycine et la proline. La glycine confère plus de flexibilité au squelette de la protéine parce qu'elle ne possède pas de chaîne latérale, tandis que la proline restreint le squelette de la protéine parce que sa chaîne latérale forme un anneau avec le squelette (Figure 2.3). Par conséquent, ces deux acides aminés se retrouvent rarement dans les hélices et les feuillets, mais ils sont souvent présents dans les coudes. La proline peut aussi être présente au début des hélices en permettant leur nucléation, car son angle dièdre $\phi \in [-60°, -77°]$ correspond aux structures hélicoïdales. Par contre, elle provoquera un noeud lorsqu'elle est dans l'hélice, car elle interrompt la séquence de ponts-H à cause de l'absence d'un hydrogène lié à l'azote dans le squelette de la proline. Les autres acides aminés, quant à eux, influencent principalement les structures tertiaire et quaternaire.

[4] La nature d'un pont-H sera présentée à la Section 2.4.

[5] En fait, le dipôle provient de l'orientation commune des charges partielles δ^- sur l'oxygène du groupe carbonyle et δ^+ sur l'hydrogène de l'amine. Ceci produit une charge résultante positive à l'extrémité N-terminal de l'hélice et une charge résultante négative à son extrémité C-terminal.

2.2.3 La structure tertiaire

La structure tertiaire est une structure principalement non-locale formée par des interactions entre des acides aminés qui sont distants dans la structure primaire de la protéine (Figure 2.2C). Ces interactions organisent les motifs de structures secondaires en supra-motifs composés uniquement de structures hélicoïdales (α), étendues (β), mixtes et séparées (α+β) ainsi que mixtes et alternées (α/β). De plus, plus une protéine est longue, plus les supra-motifs ont tendance à s'associer en différents domaines reliés entre eux par des coudes ou des boucles plus ou moins longues.

Les interactions non-locales responsables de la formation des structures tertiaires sont en ordre décroissant d'importance : les interactions non-polaire, les ponts-H, les ponts salins et les interactions van der Waals.[6] Les interactions non-polaires sont d'une importance particulièrement cruciale dans la formation du coeur de la protéine. Tous les acides aminés ne sont pas équivalents à cet égard avec les acides aminés isoleucine, phenylalanine, valine, leucine et methionine qui sont particulièrement plus non-polaires.

Les cystéines peuvent aussi jouer un rôle important dans la stabilité de la structure tertiaire lorsque deux d'entre elles forment par oxydation une liaison covalente – pont disulfure – entre le soufre de leur chaîne latérale (Figure 2.3). Cette interaction se retrouve principalement dans les protéines excrétées hors de la cellule afin de leur conférer une stabilité accrue.

2.2.4 La structure quaternaire

La structure quaternaire est la structure formée par plusieurs chaînes polypeptidiques (Figure 2.2D). Il peut s'agir de chaînes polypeptidiques identiques ou différentes qui s'auto-assemblent afin d'augmenter leur stabilité individuelle. L'arrangement final est appelé un oligomère, contrairement à un monomère lorsque la protéine possède seulement une chaîne polypeptidique. Les fonctions de plusieurs protéines dans notre cellule reposent directement sur leur structure quaternaire. Par exemple, le canal à potassium est un tétramère présent dans la membrane cellulaire qui permet le passage d'ions potassium à travers celle-ci, et l'hémoglobine est un tétramère présent dans les globules rouges qui permet le transport de l'oxygène dans le sang. Les protéines amyloïdes qui seront décrites prochainement forment, quant à elles, de petits oligomères désordonnés, mais de gros oligomères très structurés appelés fibres amyloïdes.

[6] Toutes ces interactions seront présentées à la Section 2.4.

2.3 Les méthodes expérimentales

La détermination de la structure tridimensionnelle des protéines avec une précision atomique est possible par la diffractométrie de rayons X et la résonance magnétique nucléaire [64]. Bien que n'ayant pas une résolution atomique, d'autres méthodes permettent d'obtenir des informations générales sur la structure telles que le dichroïsme circulaire [173] et la spectroscopie infrarouge à transformée de Fourier [22] pour le type de structure secondaire.

2.3.1 La diffractométrie de rayons X

L'utilité des rayons X pour déterminer la structure atomique de cristaux a été confirmée par Max von Laue[7] en 1912, soit peu après leur découverte. Deux ans plus tard, en 1914, la première structure atomique d'un cristal (NaCl) a été résolue grâce entre autres aux développements théoriques de Bragg père et fils.[8] Il a fallu cependant attendre 1937 pour que la structure atomique d'une première molécule biologique – le cholestérol – soit déterminée par Dorothy Crowfoot Hodgkin.[9] Les structures atomiques des premières macromolécules biologiques – ADN et la protéine myoglobine – ont été obtenues par la suite dans les années 1950. À ce jour, la diffractométrie de rayons X (DRX) a permis la résolution de plus de 100 000 structures atomiques de protéines, soit plus qu'avec n'importe quelle autre technique.[10]

Le principe physique de base de la DRX est la diffraction de rayons X sur un réseau d'atomes ayant des distances interatomiques similaires à la longueur d'onde des rayons X utilisés (1 Å). Afin de déterminer précisément[11] la structure atomique, les rayons X sont diffractés sur un cristal pur et très régulier composé de la protéine d'intérêt, ce qui permet d'obtenir un rapport signal sur bruit suffisant. Les rayons X diffractent sur le cristal dans des directions spécifiques, par interférence constructive, reliées à la régularité de la structure atomique du cristal. L'information contenue dans les pics de diffractions ainsi obtenus pour un ensemble complet d'orientations du

[7] Max von Laue a obtenu le prix Nobel de physique en 1914 pour sa découverte de la diffraction de rayons X par des cristaux.

[8] Sir William Henry Bragg et William Lawrence Bragg ont partagé le prix Nobel de physique en 1915 pour leurs développements théoriques dans l'analyse de la structure cristalline par diffractométrie de rayons X.

[9] D. C. Hodkin a obtenu le prix Nobel de chimie en 1964 pour la découverte de la structure atomique du cholestérol ainsi que celles de la pénicilline (1946) et de la vitamine B12 (1956).

[10] Le nombre de structures tridimensionnelles de protéines résolues par DRX qui est contenu dans la Protein Data Bank.

[11] C'est-à-dire une résolution en deçà de 2Å, car la longueur des liaisons covalentes dans une protéine sont de 1.0–2.0 Å.

cristal est ensuite utilisée pour découvrir la structure tridimensionnelle de la protéine grâce au formalisme de la transformée de Fourier.

L'application de la transformée de Fourier nécessite que l'amplitude et de la phase soient connues pour chaque pic de diffraction. L'amplitude est directement mesurée à partir de l'intensité des pics provenant des patrons de diffraction, tandis que la phase doit être déterminée en utilisant des techniques supplémentaires astucieuses.[12] Les deux techniques principalement utilisées – *Multiple isomorphous replacement* (MIR) et *Multi-* ou *single-wavelength anomalous dispersion* (MAD/SAD) – reposent sur la présence d'atomes lourds dans le cristal qui agissent comme guide dans le patron de diffraction. Plus spécifiquement, la détermination de la phase par MIR repose sur la différence entre les patrons de diffraction avec atomes lourds qui diffractent plus fortement et ceux sans, tandis que MAD/SAD repose sur l'absorption d'une partie de l'énergie des rayons X pour certaines longueurs d'onde par la présence d'atomes lourds.

Avec une amplitude et une approximation de la phase pour chaque pic de diffraction, il est possible d'appliquer la transformée de Fourier afin d'obtenir la densité électronique des atomes et, ainsi, un premier modèle atomique de la protéine.[13] Ce modèle est ensuite raffiné itérativement de façon à produire un patron de diffraction théorique en meilleur accord avec l'expérience.[14] La résolution spatiale du modèle final permet d'identifier certaines caractéristiques : la forme générale telle que la présence d'hélice alpha (résolution < 5 Å), la distinction de la forme des chaînes latérales (< 3 Å), la distinction sans ambiguïté du type de chaîne latérale (< 2 Å) et les atomes individuels (< 1 Å).

La DRX, bien qu'étant une méthode de choix, possède certaines limitations. La plus importante est l'obtention d'un cristal assez ordonné et gros de la protéine que l'on veut observer, car la croissance du cristal dépend d'un ensemble varié de conditions dont la nature de la protéine et du solvant, la présence d'ions et de ligands, la température, le pH et la cinétique de précipitation. De nos jours, ceci est facilité par l'emploi de machines qui automatisent l'essai de centaines de conditions pour la croissance des cristaux. Cependant, certaines protéines telles que les protéines

[12] La découverte des premières structures atomiques de protéines – myoglobine et hémoglobine – a été rendue possible grâce au génie de Max Perutz qui a développé la première technique – *Multiple isomorphous replacement* (MIR) – permettant de déterminer approximativement la phase.

[13] Une résolution assez précise (< 2Å) de la densité électronique permet de déterminer la position individuelle de chaque atome avec une erreur d'environ 0.1–0.2Å.

[14] La correspondance est quantifiée par le facteur de fiabilité $R = \frac{\Sigma |F_{\text{exp}} - F_{\text{modèle}}|}{\Sigma F_{\text{exp}}}$ où F(q) est le facteur de structure et les sommes sont sur tous les pics de diffraction. R doit être autour de 0.15–0.20 pour un bon modèle où $R = 0.0$ étant une correspondance parfaite.

membranaires et les petits oligomères de protéines amyloïdes sont encore difficilement cristallisables. Une autre limitation vient des dommages au cristal dus aux collisions entre les atomes et les rayons X. Ces dommages sont limités par une réduction de la température du cristal durant la DRX ainsi que par l'emploi de solutions cryoprotectrices pour éviter la formation de glace dans le cristal. Cependant, la nature même de l'environnement non-physiologique d'un cristal peut introduire des artéfacts dans la structure tridimensionnelle de la protéine. Ceci ne peut être évité, mais contrôlé par l'obtention de cristaux différents ainsi qu'en comparant les modèles DRX avec ceux obtenus par d'autres méthodes d'analyse de la structure des protéines.

2.3.2 La spectroscopie de résonance magnétique nucléaire

La spectroscopie de résonance magnétique nucléaire (RMN) est une méthode plus récente qui complète bien la DRX en permettant la résolution structurelle des protéines dans un environnement ayant des conditions physiologiques. Le phénomène de résonance magnétique nucléaire a été observé expérimentalement pour la première fois en 1938 par Isidor Isaac Rabi.[15] Le principe de base de la RMN a ensuite été perfectionné indépendamment par Felix Bloch et Edward Mills Purcell en 1946 engendrant une méthode de spectroscopie assez efficace pour étudier les propriétés magnétiques de liquides et de solides par induction magnétique.[16] Plusieurs années de développements par plusieurs chercheurs dont Kurt Wüthrich[17] ont permis la résolution de la première structure tridimensionnelle d'une protéine en 1988. À ce jour, la RMN a permis la résolution de plus de 9 000 structures atomiques de protéines, soit un nombre équivalent à environ 10% de celui résolu par DRX.

Le principe de base de la RMN est l'utilisation d'une propriété magnétique – le spin – du noyau de certains isotopes d'atomes.[18] Le spin nucléaire est relié à un moment magnétique. Ce dernier veut s'orienter parallèlement au champ magnétique externe ce qui provoque la rotation du moment magnétique autour de l'axe du champ.[19] La fréquence de rotation du moment ma-

[15] I. I. Rabi a obtenu le prix Nobel de physique en 1944 pour sa technique permettant de mesurer les propriétés magnétiques des noyaux atomiques. Il a observé qu'un champ magnétique oscillant à une fréquence particulière – fréquence de Larmor – induit une transition dans l'état du spin.

[16] F. Bloch et E. M. Purcell ont partagé le prix Nobel de physique en 1952 pour leurs développements de la RMN.

[17] K. Wüthrich a obtenu le prix Nobel de chimie en 2002 pour ces travaux permettant la détermination de la structure tridimensionnelle de macromolécules biologiques en solution par RMN.

[18] Les isotopes pertinents pour les biomolécules sont principalement ^{1}H, ^{13}C, ^{15}N et ^{31}P. Ils possèdent tous un moment cinétique de spin nucléaire de $\pm 1/2$.

[19] Ce phénomène est analogue à la précession de l'axe de rotation d'une toupie sous l'effet de la gravité.

gnétique dépend de deux facteurs : (1) de l'isotope en question, et (2) du champ magnétique senti par le noyau en prenant en compte les variations locales induites par les couches électroniques ainsi que les interactions du noyau avec les autres spins nucléaires à proximité. Cette dernière propriété est à la base de la RMN, car elle produit un changement de fréquence de rotation – appelé déplacement chimique – qui peut être mesuré afin d'extraire des informations cruciales sur l'environnement local du noyau.

Une expérience RMN est séparée en 4 phases principales : (i) application d'un champ magnétique externe assez puissant[20] afin de briser l'isotropie environnementale pour générer une magnétisation résultante de l'échantillon dans la direction du champ, (ii) envoi d'une séquence d'impulsions courtes de radio-fréquences pour changer l'orientation de la magnétisation, (iii) enregistrement en fonction du temps de la rotation et de la décroissance de cette magnétisation nouvellement induite, et (iv) analyse du signal par transformée de Fourier[21] afin d'identifier toutes les fréquences de rotation qui ont été induites dans l'échantillon ainsi que leur intensité relative.[22]

L'intérêt de la RMN pour la détermination de la structure tridimensionnelle des protéines est l'existence de protocoles d'impulsions variés qui permettent de mesurer diverses corrélations entre deux ou plusieurs spins nucléaires à proximité. Ces corrélations se produisent soit à travers l'espace ou à travers les liaisons covalentes de la protéine. Le premier type permet de déterminer les noyaux à proximité, car l'intensité de la corrélation est significative seulement pour des distances inférieures à 5 Å.[23] Le deuxième type permet de déterminer la structure locale des liaisons covalentes telle que l'angle dièdre ϕ grâce à la relation de Karplus qui relie ϕ aux valeurs de couplages scalaires mesurées : ${}^{3}\mathrm{J}_{\mathrm{HNH}\alpha} = \mathrm{A} + \mathrm{B}\ \cos(\phi) + \mathrm{C}\ \cos^2(\phi)$ où A, B et C sont des constantes déterminées empiriquement.[24] Par ailleurs, les valeurs de déplacements chimiques mesurées pour certains atomes – ${}^{1}\mathrm{H}^{\alpha}$, ${}^{13}\mathrm{C}^{\alpha}$, ${}^{13}\mathrm{C}^{\beta}$ et ${}^{13}\mathrm{C}$' – dans une protéine repliée peuvent être comparées avec leurs valeurs dans un système désordonné afin de prédire les

[20] Généralement de 1 à 20 Tesla, ce qui correspond à une fréquence de rotation de 40 à 900 MHz pour l'isotope ${}^{1}\mathrm{H}$.

[21] Technique essentielle d'analyse du signal RMN développée par Richard R. Ernst en 1966 qui lui a valu le prix Nobel de chimie en 1991.

[22] Les changements de fréquences sont comparés à une valeur de référence permettant de calculer une différence qui dépend de l'environnement local de chaque noyau. La valeur répertoriée est le déplacement chimique mesuré en partie par million.

[23] L'intensité de corrélation diminue en $\sim 1/r^6$ où r est la distance entre les deux spins.

[24] Les structures secondaires hélicoïdales (hélices-α) ont ${}^{3}\mathrm{J}_{\mathrm{HNH}\alpha} \sim 4$ Hz, tandis que les structures secondaires étendues (feuillets-β) ont ${}^{3}\mathrm{J}_{\mathrm{HNH}\alpha} \sim 9$ Hz.

régions hélicoïdales et étendues.[25]

Toutes ces informations permettent de déterminer des contraintes spatiales et géométriques sur la structure de la protéine. Ces contraintes sont ensuite utilisées dans des simulations de recuit simulé afin de trouver les structures de plus basses énergies les satisfaisant le mieux. Les modèles ainsi obtenus sont les structures RMN de la protéine. À partir de ces modèles, il est possible d'identifier les régions de la protéine qui sont structurellement très bien conservées et celles qui sont plus flexibles.

La RMN est aussi couramment utilisée pour évaluer l'accessibilité des acides aminés d'une protéine au solvant. Pour ce faire, les molécules d'eau du solvant (H_2O) sont progressivement remplacées par de l'eau lourde (D_2O) et les effets sur les spectres de déplacements chimiques sont analysés en fonction du temps. Il y a aura perte de signal pour les hydrogènes des groupes amines (H_N) du squelette de la protéine qui sont accessibles au solvant et qui ne participent pas à des ponts-H (structure secondaire) puisqu'ils s'échangeront rapidement avec les ions deutérium de l'eau lourde. Les atomes de deutérium n'apparaîtront pas dans le spectre, car ils ne seront pas stimulés par les impulsions de radio-fréquences qui stimulent l'hydrogène, leur fréquence d'oscillation/résonance étant grandement différente. Par ailleurs, la RMN permet de quantifier la dynamique d'une protéine et les interactions ligand–protéine en mesurant les temps de relaxation de la magnétisation et en mesurant les changements dans les spectres de déplacements chimiques.

Les deux limitations principales de la RMN sont la concentration de protéines dans l'échantillon qui doit être assez élevée (> 10 mM) et la taille des protéines qui ne doit pas être trop grosse (< 40 kDa). La première limitation provient de la sensibilité assez faible d'une expérience NMR à température pièce. La sensibilité peut cependant être améliorée en partie par l'utilisation d'un champ magnétique plus puissant. La deuxième limitation provient de l'augmentation du chevauchement des déplacements chimiques dans les spectres mesurés plus la protéine est grosse. De nos jours, le chevauchement peut être diminué en étiquetant sélectivement certains acides aminés de la protéine par certains isotopes et en utilisant des protocoles de pulses multi-dimensionnels.

[25] Les structures hélicoïdales sont caractérisées par une augmentation des valeurs de déplacements chimiques de $^{13}C^{\alpha}$ et $^{13}C'$ ainsi qu'une diminution de celles de $^{1}H^{\alpha}$ et $^{13}C^{\beta}$ par rapport à un système désordonné. Les structures étendues suivent la tendance contraire.

2.3.3 La spectroscopie par dichroïsme circulaire

La spectroscopie par dichroïsme circulaire (DC) est une méthode beaucoup moins précise que la DRX ou la RMN. Elle permet néanmoins d'obtenir rapidement et facilement des informations globales sur la population des structures secondaires dans les protéines [173]. Le DC est aussi une méthode de choix pour mesurer l'évolution de la structure secondaire en fonction du temps lors de processus d'agrégation complexes comme dans le cas des protéines amyloïdes.

Le principe de base du DC est que les protéines n'absorbent pas autant les rayons ultraviolets (< 260 nm) ayant une polarisation circulaire dans le sens horaire que ceux ayant une polarisation circulaire dans le sens anti-horaire. Cette différence d'absorption est influencée par la géométrie du squelette de la protéine. Le spectre d'absorption en fonction de la longueur d'onde sera donc indicatif de la structure secondaire globale de la protéine.

La clé de l'analyse d'une expérience de DC est l'utilisation d'un algorithme prédisant la probabilité globale de certaines structures secondaires – hélicoïdales, étendues (feuillets-β), coudes et désordonnées – directement à partir du spectre mesuré. Les prédictions sont préalablement calibrées et évaluées sur une banque variée de structures tridimensionnelles obtenues par DRX ayant une excellente résolution. Le spectre DC de chacune de ces protéines est mesuré dans des conditions expérimentales très similaires à celles utilisées durant la DRX. Une partie des structures est utilisée pour entraîner l'algorithme à faire les bonnes prédictions de structures secondaires à partir des spectres, tandis que l'autre partie sert à la validation de l'algorithme entraîné afin de confirmer qu'il peut faire de bonnes prédictions sur des structures tridimensionnelles différentes. Plusieurs algorithmes obtenus par des méthodes variées d'apprentissage sont disponibles pour la communauté scientifique.

La limitation fondamentale des prédictions faites à partir des spectres DC est l'influence du processus d'apprentissage utilisé pour calibrer les algorithmes de prédiction des structures secondaires à partir du spectre. En effet, la correspondance spectre–structure n'est pas unique, car deux protéines ayant des proportions très similaires de structures secondaires n'auront pas des spectres DC nécessairement très similaires. Deux algorithmes d'analyse de spectres DC appliqués sur les mêmes données ne généreront donc pas exactement le même résultat, particulièrement pour les structures étendues (feuillets-β) qui sont plus variables. Les résultats sur la proportion des coudes varieront aussi puisqu'ils ne sont pas beaucoup représentés dans les bases de structures d'entrainement utilisées durant le processus d'apprentissage. Cependant, la plupart

des algorithmes réussissent à prédire correctement le pourcentage de structures hélicoïdales, car elles sont plus régulières. L'analyse d'un spectre DC est toujours d'actualité avec le développement de nouveaux algorithmes entraînés sur des banques de structures tridimensionnelles plus complètes.

2.3.4 La spectroscopie infrarouge à transformée de Fourier

La spectroscopie infrarouge à transformée de Fourier (IRTF) est une autre méthode qui permet d'obtenir rapidement et facilement des informations globales sur la population des structures secondaires dans les protéines [22]. Comme le DC, il s'agit d'une méthode intéressante pour mesurer l'évolution de la structure secondaire en fonction du temps lors de processus d'agrégation complexes comme dans le cas des protéines amyloïdes.

Le principe de base de la IRTF est que la lumière infrarouge est absorbée par la protéine lorsque la fréquence de la lumière incidente correspond à la fréquence d'un des modes de vibrations – distance ou angle – des liaisons covalentes de la protéine. La fréquence de chaque mode de vibration dépend de la structure locale des atomes participants au mode. Le changement d'une fréquence modifie donc le spectre d'absorption, lequel est analysé pour extraire, entre autres, des informations sur les structures secondaires dans la protéine.

Les fréquences de vibrations de tous les groupements chimiques des protéines ont été mesurées expérimentalement sur des composés de référence dans l'eau (H_2O) et l'eau lourde (D_2O).[26] Le composé chimique de référence pour le squelette de la protéine est le N-methylacetamide (NMA), car il s'agit du plus petit composé ressemblant à la liaison peptidique entre deux acides aminés consécutifs. Ce composé possède 6 modes principaux de vibration qui ont été mesurés expérimentalement et expliqués théoriquement par des calculs *ab initio*. Celui utilisé pour la détection de structures secondaires est le mode Amide I, car il est très sensible à la structure du squelette de la protéine et très peu à celle de la chaîne latérale. Ce mode possède une fréquence de ~ 1650 cm^{-1} et consiste en la vibration de la liaison C=O (70–85%) couplée à celle de la liaison C–N (10–20%).

Dans le contexte d'une protéine, l'expérience et les calculs *ab initio* montrent que la fré-

[26] Il est souvent nécessaire de comparer le spectre d'absorption d'une protéine dans l'eau avec celui obtenu dans l'eau lourde, car les fréquences de vibrations – et donc d'absorption – ne seront pas les mêmes. Ceci permet d'éliminer la présence de certains chevauchements dans le spectre d'absorption et de confirmer les prédictions structurelles puisqu'elles ne devraient pas dépendre du type d'eau utilisé.

quence du mode Amide I change selon la présence de structures secondaires : hélice-α (1648–1657 cm^{-1}), feuillets-β (1623–1642 cm^{-1} / 1674–1695 cm^{-1}), coudes (1662–1686 cm^{-1}) et désordonnées (1642–1657 cm^{-1}).[27] L'apparition de nouvelles bandes d'absorption ainsi que leur intensité à ces fréquences indiquent la présence des structures secondaires correspondantes. L'analyse du spectre IRTF peut se faire grâce à des algorithmes entraînés, en suivant un protocole similaire à celui de la DC, pour identifier les structures secondaires globales directement à partir du spectre. Dans le cas où seulement l'identification de la présence de motifs est requise, le spectre peut être traité afin de faire ressortir les fréquences d'absorption permettant d'identifier la présence de ces motifs structuraux.

La limitation principale de la IRTF est très similaire à celle du DC : les algorithmes d'analyse du spectre ne font pas exactement les mêmes prédictions, car la correspondance spectre–structure n'est pas unique et la prédiction dépend ultimement du protocole d'apprentissage utilisé pour calibrer l'algorithme. Dans le cas de la IRTF, il y a aussi des absorptions proches de la bande Amide I qui proviennent à 10–30% des chaînes latérales ce qui ajoute une variabilité au spectre. Contrairement au DC, la IRTF est très efficace pour détecter la présence de feuillets-β, mais elle l'est moins pour détecter la présence d'hélices-α. Par ailleurs, la IRTF peut aussi être utilisée pour quantifier l'accessibilité de la protéine au solvant par échange hydrogène/deutérium ($^1H/^2H$) en remplaçant l'eau (H_2O) par de l'eau lourde (D_2O). L'information ainsi récoltée n'est cependant pas spécifique aux acides aminés contrairement à la RMN.

2.4 Les interactions physiques dans les protéines

Les protéines sont des molécules nanoscopiques qui obéissent donc aux lois de la physique quantique. Les phénomènes de repliement et d'oligomérisation des protéines font cependant intervenir des échelles temporelles et des changements structuraux pour lesquels les phénomènes quantiques peuvent être traités classiquement par des modèles empiriques.[28] De tels modèles sont présentés en détail à la Section 3.1.

Les interactions dans les protéines peuvent être séparées en deux types : les interactions

[27] Les valeurs données correspondent à une solution de H_2O et elles sont différentes dans le cas d'une solution de D_2O.

[28] La nature quantique des protéines doit cependant être prise en compte lorsqu'il y a bris ou formation de liaisons covalentes lors de l'activité enzymatique, par exemple. Une quantification précise des interactions protéine–ion ainsi que l'absorption et l'émission de photons par une protéine requièrent aussi le formalisme de la mécanique quantique.

liées – angles de valence et angles dièdres – qui font intervenir des liaisons covalentes ainsi que les interactions non-liées – van der Waals, électrostatiques, ponts-H et non-polaires – qui agissent à travers l'espace. L'origine quantique de ces interactions et leur équivalence classique sont abordées dans cette section [135, 191, 410].

2.4.1 Les interactions liées

Les interactions liées permettent de maintenir l'intégrité de la structure atomique des acides aminés de la protéine (Figure 2.3). Elles font principalement intervenir une, deux ou trois liaisons covalentes consécutives (Figure 2.6).

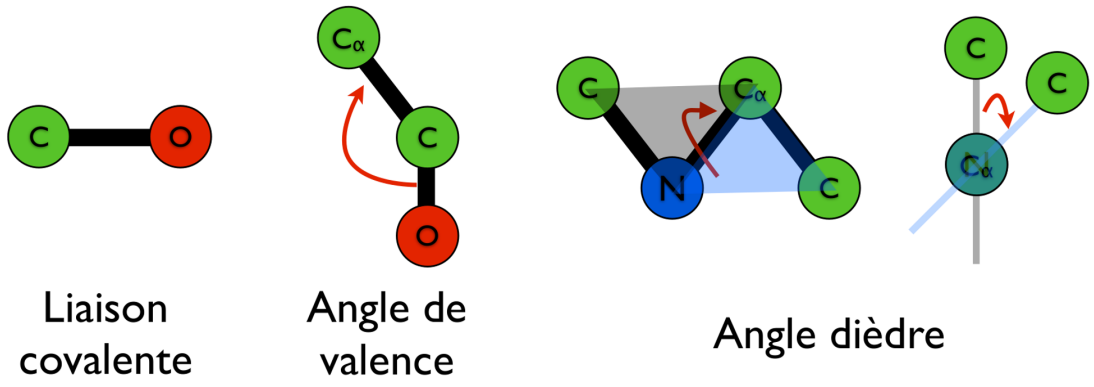


Figure 2.6 – Schématisation des interactions liées dans les protéines.

Une liaison covalente est formée entre deux atomes voisins par un partage d'électrons provenant de leur couche de valence incomplète afin de la remplir. Le nombre d'électrons disponibles par atome pour ce partage correspond au nombre présent dans sa couche de valence.[29] Un atome peut donc partager plusieurs électrons avec un seul autre atome – auquel cas l'énergie d'association croît avec le nombre d'électrons participants à la liaison covalente – ou il peut les partager avec un maximum de quatre autres atomes.

Lorsqu'un atome participe à plusieurs liaisons covalentes, celles-ci adoptent des orientations relatives par rapport à l'atome central. L'angle entre deux liaisons covalentes séparés par un atome central est appelé l'angle de valence. Il y a une ensemble très restreint de géométries possibles, car les orientations dépendent du type d'hybridation des orbitales électroniques de

[29] La couche de valence correspond au dernier niveau d'énergie qui n'est pas rempli complètement d'électrons ; les autres niveaux de plus basse énergie l'étant.

l'atome central. Dans le cas des protéines, la géométrie locale des atomes est trigonale plane ou tétraèdre dont les angles entre les liaisons covalentes par rapport à l'atome central sont respectivement de $\sim 120°$ et $\sim 109.5°$.[30] Lorsque les composés liés à l'atome central sont différents, de légères variations par rapport à ces valeurs idéales d'angles sont introduites.

Les densités électroniques de deux atomes espacés par trois liaisons covalentes interagissent aussi ensemble. Les quatre atomes impliqués se positionnent préférentiellement selon certaines orientations entre les deux plans formés respectivement des liaisons covalentes 1 et 2, et 2 et 3. L'angle entre ces deux plans est appelé l'angle dièdre. L'origine de cette interaction est principalement associée à la répulsion entre les couches électroniques des atomes : les orientations stables minimisent cette répulsion. Un phénomène de résonance entre les densités électroniques permet aussi de stabiliser ces orientations particulières.

À température pièce, l'énergie des vibrations thermiques est insuffisante pour briser les liaisons covalentes et, dans une première approximation, leur longueur oscille autour d'une valeur d'équilibre tel un oscillateur harmonique. Les angles dièdres oscillent aussi harmoniquement autour d'une valeur d'équilibre. Les angles dièdres adoptent principalement certaines valeurs bien que toutes les valeurs soient accessibles. Les paramètres de ces interactions sont obtenus expérimentalement ainsi que par calcul *ab initio*. Les modèles classiques et leurs paramètres sont abordés en détail à la Section 3.1.

2.4.2 Les interactions non-liées

Les interactions non-liées sont responsables des structures secondaire, tertiaire et quaternaire des protéines. Les quatre types qui dominent sont les interactions de van der Waals, électrostatiques, de ponts hydrogènes et non-polaires telles qu'illustrées à la Figure 2.7. Contrairement aux interactions liées, elles ne font pas intervenir de liaisons covalentes.

Les interactions de van der Waals rendent compte de la déviation observée dans la pression d'un gaz par rapport à la pression théorique d'un gaz idéal.[31] En effet, même le comportement d'un gaz monoatomique noble – par conséquent non-chargé et non-réactif – dévie du gaz idéal dû

[30] La géométrie dépend du type d'orbitales hybridées formées : orbitales *sp* pour linéaire (180°), orbitales sp^2 pour trigonale plane (120°) et orbitales sp^3 pour tétraèdre (109.5°).

[31] Pour des densités faibles, l'équation de van der Waals relie la pression (p) à la température (T), au volume (V) et au nombre de particules (N) selon $p = \frac{NkT}{V-Nb} - \frac{aN^2}{V^2}$ où a et b sont des paramètres qui dépendent du détail des interactions interatomiques. Pour des densités encore plus faibles, cette équation devient la relation des gaz parfaits : $pV = NkT$ qui est indépendante des paramètres a et b.

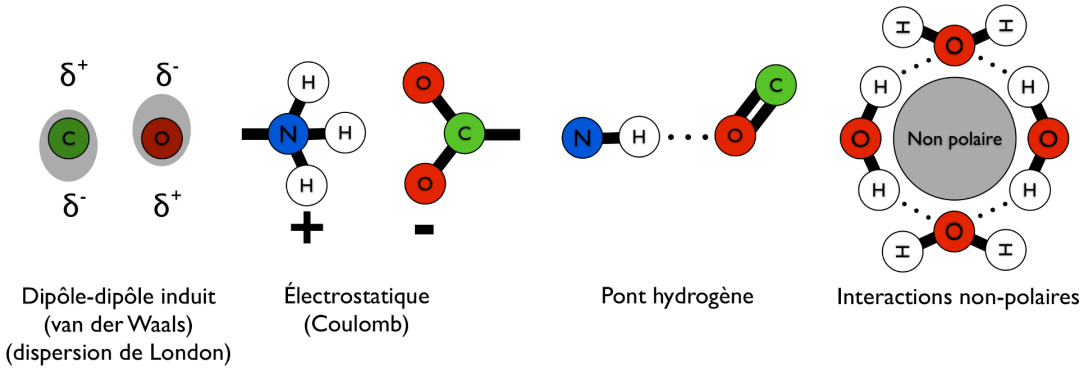


Figure 2.7 – Schématisation des interactions non-liées dans les protéines. Les charges partielles sont illustrées par δ^+ et δ^- et les ponts-H par les traits noirs pointillés.

à la présence d'interactions interatomiques. L'origine de ces interactions attractives entre deux atomes est due à la formation de dipôles induits par des fluctuations de la densité électronique autour du noyau de chaque atome. Ce phénomène est nommé la force de dispersion de London. Il est significatif seulement à très courte distance entre les deux atomes (< 1 nm). L'attraction est maximale à une distance d'environ 0,3 nm qui dépend du type de la paire d'atomes. À plus courte distance, l'interaction devient peu à peu dominée par la répulsion entre les couches électroniques des deux atomes due au principe d'exclusion de Pauli. Classiquement, il s'agit d'une interaction entre le champ électrique émis par le dipôle d'un atome et un dipôle induit dans un autre atome par ce même champ.

Les interactions électrostatiques proviennent des charges partielles portées par chaque atome ainsi que des charges ioniques. Ce sont des interactions de longues portées qui sont décrites par le potentiel Coulomb. Les charges partielles découlent d'un partage inégal de la densité électronique autour de deux atomes participant à une liaison covalente. Les atomes électronégatifs ont tendance à attirer plus la densité électronique de la liaison covalente vers eux. La valeur d'une charge partielle dépend donc du type des atomes participants aux liaisons covalentes. Elles sont aussi sujettes à des fluctuations dues à leur environnement local qui peut changer. Dans les modèles classiques, les charges partielles sont habituellement fixes ou basées sur des modèles simples de polarisation.

Les interactions pont hydrogène (pont-H) se produisent entre deux groupes d'atomes : (i) un groupe donneur consistant en un hydrogène qui partage une liaison covalente avec un atome très électronégatif (azote, oxygène ou souffre), représenté par X–DON–H, et (ii) un groupe accep-

teur consistant en un atome très électronégatif qui partage une liaison covalente avec un atome qui l'est beaucoup moins (carbone), représenté par Y–ACC. Par conséquent, l'hydrogène du groupe donneur et l'atome électronégatif du groupe accepteur possèderont respectivement une charge partielle positive et négative. Dans le squelette de la protéine, par exemple, la présence de ponts-H entre les groupements X–N–H^{δ^+} et Y–C=O^{δ^-} stabilise la formation des structures secondaires (Figure 2.5). Le premier groupe agit, en effet, comme un donneur d'hydrogène (H^{δ^+}) au deuxième groupe dont l'oxygène agit comme un accepteur (O^{δ^-}) engendrant une interaction ayant un caractère partiellement dipôle-dipôle et covalent. Il s'agit d'une interaction de très courte portée, avec une distance optimale de 0,18 nm, qui est plus forte qu'une interaction van der Waals.[32] Une autre caractéristique fondamentale des ponts-H est que leur force dépend de l'orientation respective des groupes donneur et accepteur : l'interaction décroît lorsque l'angle entre le triplet ACC$\cdots$H–DON dévie de sa valeur idéale de 180°. Il est commun de ne pas considérer un pont-H comme étant formé lorsque la déviation est de plus de 30°. Classiquement, une combinaison adéquate des interactions de van der Waals et électrostatiques entre les groupes donneur et accepteur semble suffisante pour décrire la formation des ponts-H malgré l'absence du caractère anisotrope de cette interaction.

Les interactions non-polaires sont fondamentales dans la structure et la fonction des protéines bien qu'elles ne soient pas des interactions à proprement parler. Elles proviennent indirectement des interactions entre la protéine et le milieu aqueux environnant. Grâce à leur polarité, les molécules d'eau jouent un rôle essentiel dans la stabilité des protéines : (i) elles interagissent favorablement avec les groupements polaires de la protéine et (ii) elles isolent du solvant les groupements non-polaires de la protéine avec lesquels elles ne peuvent pas interagir. C'est ce dernier phénomène qui est appelé interactions non-polaires, car les groupements non-polaires de la protéine semblent interagir entre eux en se regroupant. C'est le même phénomène qui est à l'origine de la non-miscibilité de l'eau et de l'huile (formée de composés non-polaires). À sa source, les molécules d'eau (polaires) s'ordonnent autour des composés non-polaires afin de maximiser leurs interactions avec d'autres molécules d'eau puisqu'elles interagissent faiblement avec les composés non-polaires. Ceci crée une tension de surface non-favorable autour des composés non-polaires qui est réduite en diminuant l'aire de la surface de contact entre les molécules d'eau et les composés non-polaires. Dans le cas des protéines qui possèdent des groupements

[32]Cette distance optimale est près de 2 fois plus petite que celle pour une interaction entre deux dipôles induits (van der Waals).

polaires et non-polaires, ces derniers seront donc isolés le plus possible du solvant aqueux pour maximiser les interactions entre les groupements polaires et les molécules d'eau.

Comme discuté dans la prochaine section, toutes ces interactions, particulièrement celles non-polaires, sont cruciales dans le repliement – formation de structures tridimensionnelles – des protéines. Leur modélisation mathématique est présentée plus loin à la Section 3.1.

2.5 Le repliement des protéines

En présence de conditions physiologiques[33] adéquates, les protéines se replient en une structure tridimensionnelle (structures secondaire, tertiaire et quaternaire) caractéristique de leur séquence d'acides aminés (structure primaire) tel qu'illustré précédemment à la Figure 2.2. Chaque type de protéine – globulaire, fibrillaire, membranaire et désordonnée – se replie différemment dû à des environnements fonctionnels et à des ensembles d'interactions globalement différents.

Plusieurs études théoriques, expérimentales et de simulations se sont attardées à décrire qualitativement et quantitativement les contributions principales au repliement des protéines globulaires. Ceci a permis le développement d'un modèle de repliement général qui est globalement accepté, mais dont les détails sont toujours sujets à discussion [136, 224, 238, 507].

Les protéines intrinsèquement désordonnées sont de plus en plus reconnues comme essentielles au fonctionnement de la cellule [353, 477, 508]. Elles sont cruciales dans la régulation cellulaire, car elles adoptent une structure tridimensionnelle donnée seulement sous des conditions particulières telles que la présence de leurs protéines d'interaction. Les protéines amyloïdes sont principalement désordonnées au niveau du monomère, mais elles s'auto-assemblent en structures ordonnées appelées fibres amyloïdes par un processus de nucléation [339]. Elles sont souvent associées à des maladies neurodégénératives. Les protéines amyloïdes peuvent aussi interagir avec les membranes de phospholipides et y former, entre autres, des pores [263, 264, 525]. Les chemins de repliement et d'agrégation des protéines amyloïdes sont donc assez complexes tels qu'illustrés à la Figure 2.8.

L'exemple le plus fameux et le plus ancien est la protéine bêta-amyloïde associée à la maladie d'Alzheimer qui sera étudiée aux Chapitres 5, 6 et 7. La protéine Huntingtine est une autre protéine amyloïde qui sera étudiée aux Chapitres 9, 10, 11 et 12.

[33] Condition semblable à l'environnent de la protéine dans la cellule.

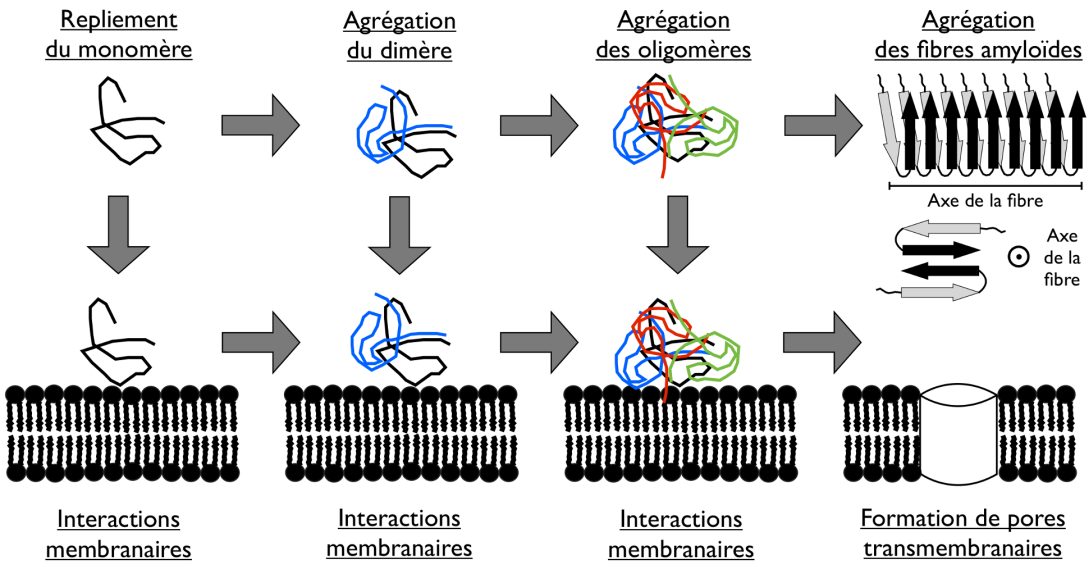


Figure 2.8 – Les chemins du repliement des protéines amyloïdes. La structure de la fibre amyloïde illustrée est l'une de celles de la protéine bêta-amyloïde (Aβ) [470] ; cette structure dépend, bien que son coeur soit toujours riche en feuillets-β, de la protéine amyloïde en question et des conditions de croissance de la fibre.

2.6 Les interactions protéine–membrane lipidique

Les membranes lipidiques jouent un rôle crucial dans le fonctionnement de la cellule. Cette dernière est, en fait, entourée d'une membrane lipidique qui isole l'intérieur de la cellule de son environnement extérieur telle qu'illustrée précédemment à la Figure 2.1. La membrane sert aussi de support et de régulateur pour un ensemble de protéines membranaires qui contrôlent la communication et les échanges de composés entre la cellule et son milieu environnant. À l'intérieur d'une cellule eucaryote[34], le noyau et les organites – les mitochondries, le réticulum endoplasmique, l'appareil de Golgi et les lysosomes – sont entourés d'une membrane lipidique qui permet de séparer et de contrôler leur environnement interne indépendamment du reste de la cellule.

Dans cette section, l'origine de la stabilité de ces membranes lipidiques est présentée. Les interactions entre une protéine et une membrane sont aussi abordées. Aux Chapitres 11 et 12, les interactions huntingtine–membrane sont étudiées à l'aide de simulations numériques.

[34]Type de cellule à la base des organismes pluricellulaires animal et végétal.

2.6.1 La structure des membranes lipidiques

La membrane lipidique est présente dans le contexte cellulaire en tant que membrane autour de la cellule, du noyau et des organites (Figure 2.9) [313]. Auparavant, la membrane était décrite par le modèle de la mosaïque fluide qui rend compte de sa nature fluide, de son rôle de support pour les protéines membranaires ainsi que de son rôle de séparateur entre deux milieux [433]. Il est maintenant reconnu que la membrane cellulaire est un système beaucoup plus complexe. Par exemple, la membrane lipidique se déforme à proximité des protéines membranaires, certains lipides interagissent de façon privilégiée avec certaines protéines, et les protéines membranaires ainsi que les lipides ne diffusent pas librement [147]. L'activité de certaines protéines membranaires peut aussi être régulée par la membrane lipidique [378]. Par ailleurs, la présence de rafts lipidiques, lesquels auraient un rôle important dans la compartimentation des protéines membranaires et dans la signalisation cellulaire [58, 282], changent localement les propriétés de la membrane [153, 472]. Ceux-ci jouent aussi un rôle important lors des processus de fusion cellulaire [431]. Cette liste ne montre qu'une partie des rôles complexes de la membrane cellulaire et elle s'accroît continuellement.

Les phospholipides sont le constituant principal de la plupart des membranes cellulaires (Figure 2.10) [313, 502]. Ils sont des molécules ayant un caractère amphiphile puisqu'ils possèdent une tête polaire avec un groupement phosphate, et une queue non-polaire avec deux chaînes acyles composées de groupements CH_2. Il existe plusieurs types de phospholipide, lesquels sont classifiés selon la composition moléculaire de leur tête polaire et de leur queue non-polaire telles que montrées à la Figure 2.10 [182]. La queue non-polaire possède une variabilité quant à la longueur des chaînes acyles et au nombre d'insaturations, c'est-à-dire de liaisons covalentes doubles entre les atomes de carbone. De plus, certains phospholipides, appelés les lysophospholipides, possèdent seulement une chaîne acyle. Par ailleurs, la tête polaire peut avoir diverses tailles et charges totales.

Puisque chaque type de phospholipide a une structure moléculaire différente, chacun influence les propriétés membranaires différemment (Figure 2.11). Par exemple, les insaturations causent une diminution de la flexibilité des chaînes acyles, ce qui réduit l'épaisseur ainsi que la température de transition de la phase gel à la phase fluide des membranes lipidiques [271]. À l'opposé, des chaînes acyles plus longues sont corrélées avec une augmentation de cette température de transition. La structure moléculaire de la tête polaire a aussi des conséquences

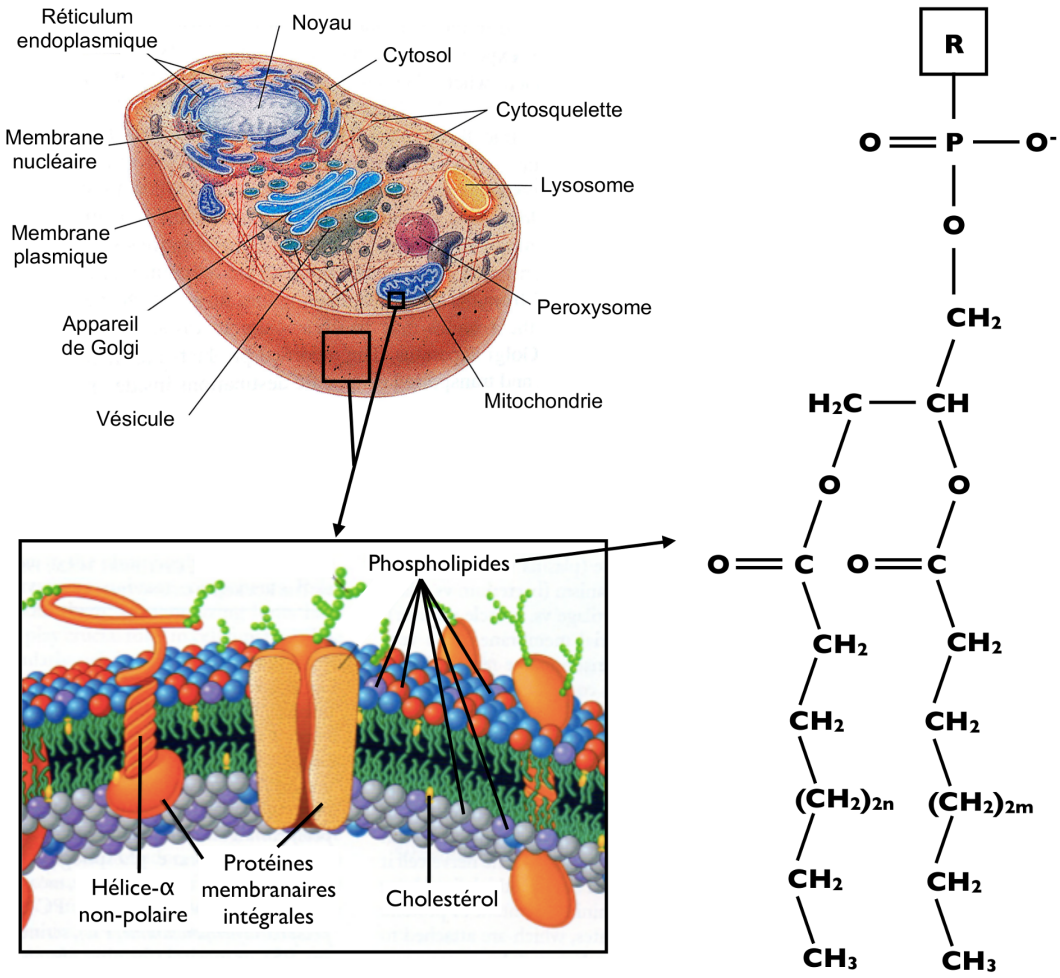
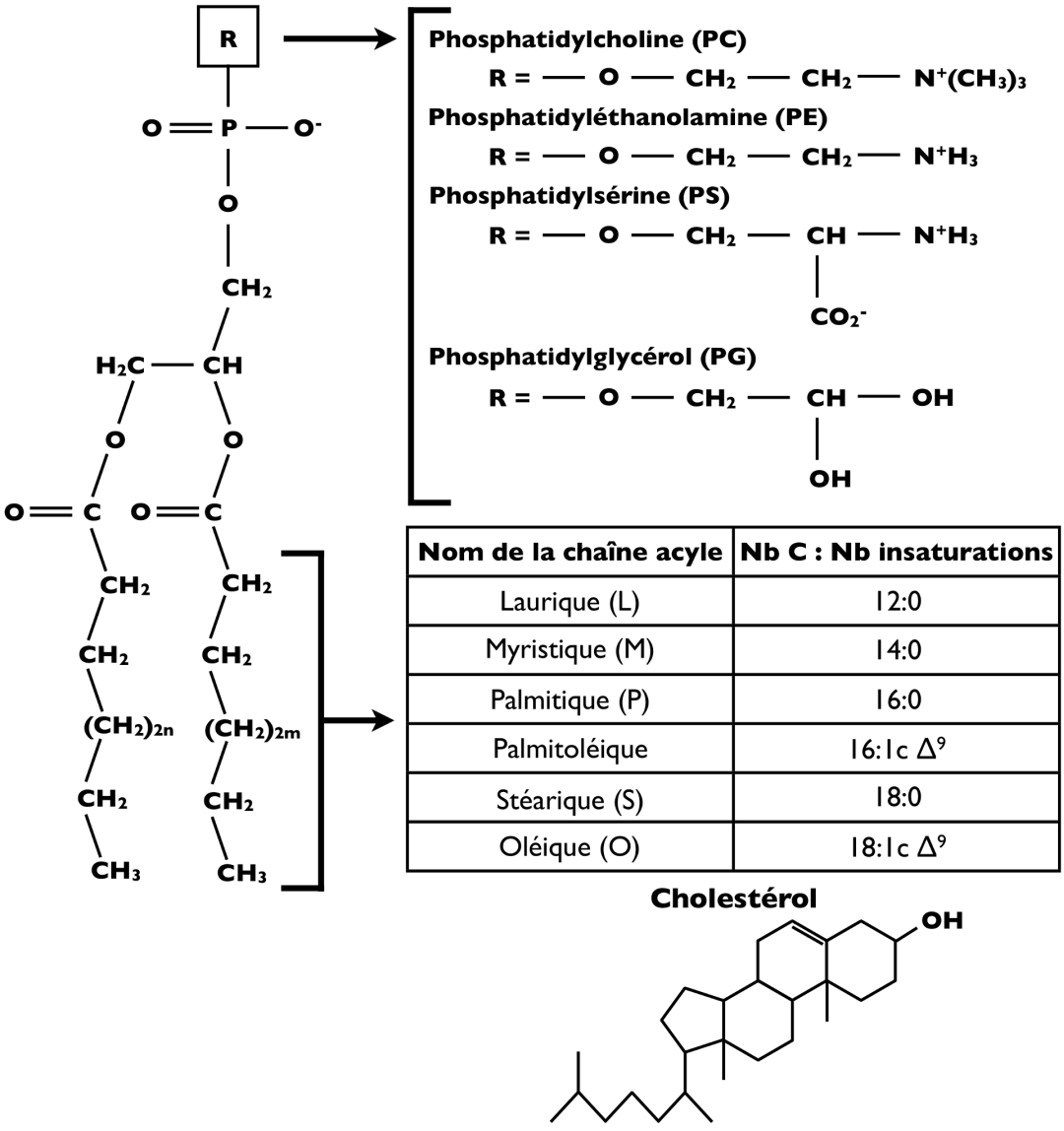


Figure 2.9 – La membrane cellulaire et les membranes des organites dans la cellule sont composées de phospholipides. Image adaptée [182, 191].

sur les transitions de phases des membranes lipidiques [271]. La forme globale des phospholipides génère quant à elle une courbure spontanée dans les membranes lipidiques composées de certains types de phospholipide tels que les phosphatidyléthanolamines (PE) et les lysophosphatidylcholines (lyso-PC) (Figure 2.11) [174, 199]. Par ailleurs, la présence d'autres types de lipide tels que les sphingomyélines et le cholestérol influencent aussi les propriétés membranaires [282, 364, 393, 514]. Chaque type de cellule et d'organite possède une composition unique de lipides [502]. Cependant, à ce jour, il reste beaucoup à comprendre concernant le rôle des proportions précises de lipides observées physiologiquement.

Dans un milieu aqueux, les phospholipides s'agrègent en structures ordonnées. Ce phéno-



Nom de la chaîne acyle	Nb C : Nb insaturations
Laurique (L)	12:0
Myristique (M)	14:0
Palmitique (P)	16:0
Palmitoléique	16:1c Δ^9
Stéarique (S)	18:0
Oléique (O)	18:1c Δ^9

Figure 2.10 – Les différents types de phospholipide ainsi que la structure du cholestérol.

mène est principalement conditionné par la ségrégation des chaînes non-polaires des molécules d'eau. La tête polaire, quant à elle, reste en contact avec le solvant aqueux. Il y a agrégation lorsque la densité de lipide dépasse une valeur critique qui dépend du nombre de carbone dans les chaînes acyles [57]. Les morphologies résultantes dépendent principalement de la courbure intrinsèque des phospholipides qui s'agrègent (Figure 2.11) [174, 199]. Par exemple, les phases micelle, hexagonale et lamellaire sont particulièrement accessibles à lyso-PC, PE et PC, respectivement. D'autres morphologies telles que les vésicules multi-lamellaires et unilamallaires, les

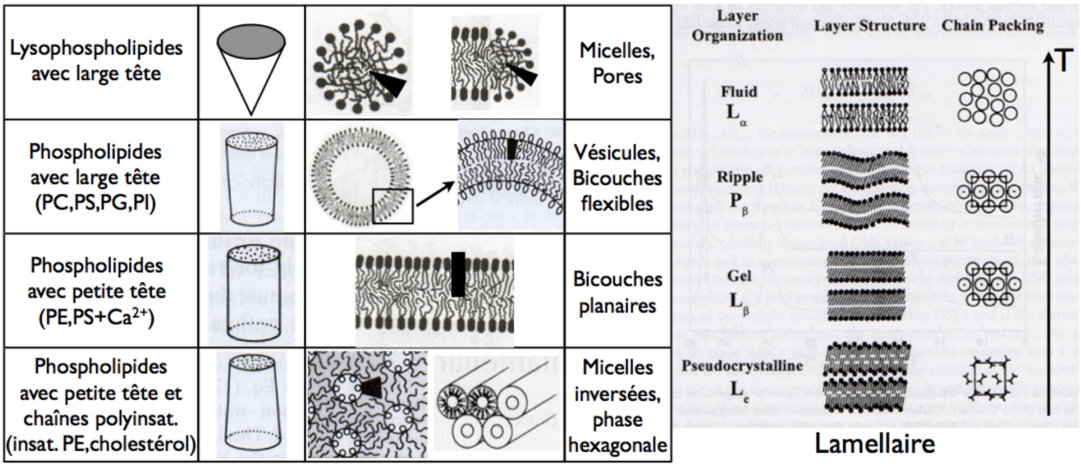


Figure 2.11 – Exemples de propriétés de la membrane influencées par le type de phospholipide. À gauche, la courbure de la membrane dépend de la forme effective des phospholipides. À droite, les différentes phases de la membrane lamellaire où la phase fluide est l'état principalement observé physiologiquement. Image adaptée [174, 199].

bicelles, et les phases cubiques sont aussi observées.

Dans les Chapitres 11 et 12, les interactions huntingtine–membrane seront analysées en présence d'une membrane lamellaire puisque la membrane peut être considérée plane, dans une première approximation, à l'échelle des interactions peptide–membrane qui sont de l'ordre du nanomètre. La phase lamellaire est une bicouche plane composée de phospholipides dont l'arrangement moléculaire dépend de la température (Figure 2.11) [271]. Seulement la phase fluide sera considérée puisqu'il s'agit de la phase majoritairement observée physiologiquement et utilisée dans les expériences sur les interactions peptide–membrane. Il est important de noter que ces modèles ont cependant leurs limites puisque la membrane cellulaire possède une distribution hétérogène de lipides dont la phase et les propriétés structurales peuvent varier [153, 472]. Les rafts lipidiques qui sont composés de molécules de cholestérol et de sphingomyéline sont un exemple [58, 282].

2.6.2 Les interactions protéine–membrane

Les peptides membranaires interagissent favorablement avec les membranes lipidiques dû à leur caractère amphiphile, c'est-à-dire à la fois polaire et non-polaire [32, 165, 505]. Cette caractéristique est retrouvée, par exemple, chez les peptides antimicrobiens, qui fournissent une dé-

fense contre les bactéries chez divers organismes, et chez les ancres peptidiques, qui sont un segment d'une protéine contrôlant sa localisation sur certains types de membrane dans la cellule. L'insertion du peptide[35] est caractérisée par quatre étapes : (i) l'adsorption du peptide, (ii) son repliement à la surface de la membrane lipidique, (iii) son insertion dans la membrane et, dans certains cas, (iv) son auto-assemblage avec d'autres peptides insérés (Figure 2.12) [505]. Chacune de ces étapes fait intervenir des parties spécifiques de la membrane lipidique, laquelle est décrite par trois régions aux propriétés différentes : le centre non-polaire, la surface de la membrane (région mitoyenne) et l'extérieur très polaire.

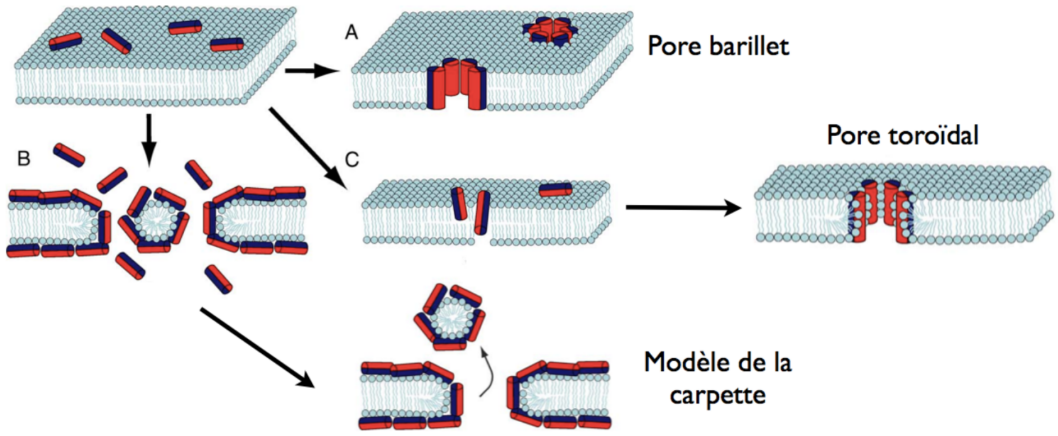


Figure 2.12 – Repliement et agrégation de peptides amphiphiles sur une membrane de phospholipides. Les côtés non-polaire et polaire des peptides sont respectivement colorés en bleu et en rouge. Image adaptée [83].

La première étape est favorisée par la présence d'interactions électrostatiques longues portées. La deuxième étape est caractérisée par un changement structurel significatif du peptide lorsqu'il passe du solvant aqueux à la surface de la membrane. Cette réorganisation permet aux acides aminés non-polaires du peptide de pointer vers l'intérieur de la membrane lipidique, tandis que les acides aminés polaires pointent vers le solvant [32, 505]. Certains acides aminés aromatiques comme la tryptophane servent d'ancre dans la membrane lipidique, tandis que des acides aminés chargés positivement comme l'arginine interagissent préférablement avec la tête polaire des phospholipides [83]. À la surface, le peptide se replie en hélice-α la plupart du temps, ce qui permet de neutraliser les charges partielles de son squelette polypeptidique

[35] Pour simplifier la discussion, le terme peptide est employé autant pour les peptides antimicrobiens que pour les ancres membranaires.

grâce à des ponts hydrogènes entre les groupements C=O et N–H. Ces derniers doivent être neutralisés avant l'étape d'insertion puisque la présence de groupements polaires dans le centre non-polaire de la membrane est énergétiquement très défavorable. La troisième étape est caractérisée par l'insertion du peptide dans la membrane, laquelle peut être favoriser par la présence de motifs dans l'ensemble structurel du peptide en solvant. Le peptide inséré possède une orientation qui est soit parallèle ou perpendiculaire (transmembranaire) à la surface de la membrane. L'orientation parallèle se produit pour les peptides amphiphiles. Elle est caractérisée par un alignement du plan amphipathique[36] du peptide avec l'interface polaire/non-polaire des phospholipides de la membrane et par une orientation des acides aminés non-polaires vers le centre de la membrane. L'insertion perpendiculaire se produit pour des peptides principalement non-polaires ou pour des peptides amphiphiles seulement lorsqu'ils s'auto-assemblent en pores (quatrième étape) [83, 194, 505].

[36]Les acides aminés non-polaires et polaires se retrouvent, grosso-modo, de part et d'autre du plan amphipathique du peptide. Ce plan contient l'axe de l'hélice alpha du peptide.

CHAPITRE 3

LES MÉTHODOLOGIES D'EXPÉRIENCES *IN SILICO*

L'émergence du domaine de la biologie moléculaire avec la découverte des premières structures tridimensionnelles de biomolécules dans la deuxième moitié du 20^{e} siècle s'est produite en parallèle avec l'avènement des ordinateurs modernes. L'intérêt d'appliquer les ordinateurs pour l'étude des structures et des fonctions des protéines a été rapidement réalisé dans les années 60. Les expériences *in silico* font, en effet, le pont entre la théorie et l'expérience en laboratoire grâce à des simulations de modèles physiques par ordinateurs afin de décrire les mécanismes dans les protéines avec une résolution atomique (Figure 3.1).

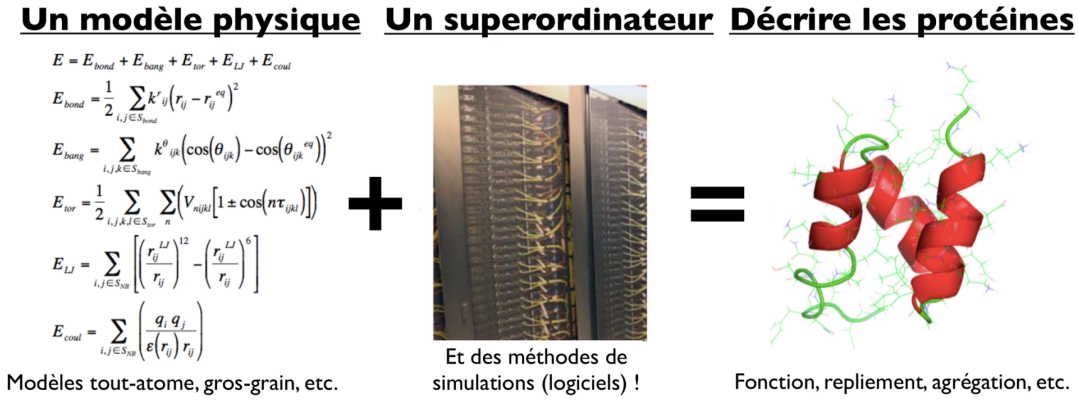


Figure 3.1 – Schématisation d'un protocole de simulation de protéines.

Les protéines appartiennent au monde nanoscopique décrit par la mécanique quantique (MQ) tel que présenté à la Section 2.4. Les changements structurels se produisant dans une protéine lors du repliement et de l'auto-assemblage se passent, néanmoins, sur des échelles de temps et d'espace pouvant être étudiées convenablement par la mécanique moléculaire (MM), c'est-à-dire par l'utilisation de modèles empiriques. Ces approximations sont cruciales, car l'étude de ces mécanismes par la MQ est tout simplement hors de portée de la puissance de calculs des ordinateurs pour les années à venir.[1]

[1]En effet, simuler 50 molécules d'eau en utilisant la MQ – simulations de Car-Parrinello – pour 2 ns prend plusieurs jours sur un superordinateur [464], tandis que la même simulation avec un potentiel MM empirique prend quelques minutes sur un ordinateur de bureau. Il est aussi possible de combiner les deux formalismes dans des simulations hybrides MQ/MM afin de décrire certaines zones d'intérêt – telles que les sites enzymatiques et les

Le premier modèle MM – aussi communément appelé un potentiel ou un champ de force – décrivant les énergies et les forces internes dans une molécule au niveau atomique a été développé en 1967 par Shneior Lifson et Arieh Warshel [276]. L'année suivante Michael Levitt et Shneior Lifson ont appliqué un modèle similaire pour le raffinement des structures atomiques tridimensionnelles de protéines par minimisation d'énergie [268]. C'est littéralement la naissance d'un nouveau champ de recherche : la biophysique computationnelle. Un peu moins de dix années plus tard, J. Andrew McCammon, Bruce R. Gelin et Martin Karplus ont simulé une protéine (BPTI), pour la première fois, par dynamique moléculaire en 1977 [312]. Une autre décennie plus tard, avec l'avènement d'ordinateurs plus puissants, la première dynamique moléculaire d'une protéine avec une représentation explicite des molécules d'eau du solvant a été réalisée par Michael Levitt et Ruth Sharon en 1988 [269].

Durant les dernières années, les progrès fulgurants de la puissance de calculs des ordinateurs et de l'efficacité des algorithmes de dynamique moléculaire ont sans cesse repoussé les limites de taille et de durée des systèmes moléculaires étudiés par simulation [65, 480]. Le superordinateur ANTON, par exemple, est spécifiquement conçu pour exécuter des dynamiques moléculaires de protéines à la milliseconde en quelques semaines [280, 425–427].[2] Le projet Folding@home, de son côté, utilise la puissance de calculs de milliers d'ordinateurs personnels délocalisés et des algorithmes novateurs de reconstruction de la dynamique afin d'accéder à des échelles de temps similaires [29, 262, 365]. L'utilisation de superordinateurs ayant toujours plus de processeurs permet la simulation de systèmes ayant plus d'un million d'atomes [480]. La biophysique computationnelle est un domaine de recherche mature, partenaire à part entière des expériences en laboratoire [411] en permettant de caractériser la cinématique et la thermodynamique des protéines à des échelles de temps et d'espace difficilement accessibles expérimentalement [139].[3]

Malgré ces incroyables avancements technologiques, il restera toujours des systèmes et des mécanismes plus complexes en terme d'échelles de temps et de grandeur. Deux autres avenues principales sont empruntées pour accélérer les simulations : (i) la réduction de la complexité

sites d'interactions ion–protéine – précisément avec la MQ tout en gardant des temps de calculs rapides en traitant le reste des atomes avec la MM. Le développement de méthodes multi-échelles comme la MQ/MM est à l'origine du prix Nobel de chimie octroyé à Martin Karplus, Michael Levitt et Arieh Warshel en 2013.

[2] ANTON est un superordinateur développé par D. E. Shaw Research qui exécute un logiciel de dynamique moléculaire (Desmond) spécialement conçu et optimisé pour ANTON [425, 426]. Dans ses domaines d'application, ANTON est plus de 100 fois plus rapide que les superordinateurs conventionnels.

[3] Le superordinateur peut être littéralement vu comme un « Computational microscope » – Microscope computationnel – de la biologie moléculaire.

du modèle simulé en regroupant, par exemple, les atomes en pseudo-atomes afin de créer de nouveaux modèles dits gros-grains,[4] et (ii) l'utilisation d'algorithmes d'échantillonnage d'événements plus performants que la dynamique moléculaire. La première stratégie agit directement sur la rapidité de calcul : moins d'atomes implique moins d'interactions à calculer par l'ordinateur. La deuxième stratégie repose sur l'utilisation plus judicieusement les ressources informatiques afin de quantifier un phénomène donné plus rapidement.

Dans ce chapitre, les deux éléments essentiels d'une simulation numérique de protéines sont abordés : (i) un modèle général qui décrit les interactions en présence, et (ii) une méthode de simulation qui utilise le modèle pour quantifier des événements concrets. Plusieurs types de modèles atomiques et de méthodes de simulation existent pour divers ensembles de problématiques à résoudre [2]. Ainsi, l'emphase de ce chapitre porte sur les approches utilisées dans les prochains chapitres.

3.1 Les modèles empiriques

Les modèles empiriques de la mécanique moléculaire (MM) – aussi appelés champ de force ou potentiel – décrivent approximativement les interactions physiques interatomiques. Tous les champs de force modernes qui sont exprimés en coordonnées cartésiennes pour les protéines, les acides nucléiques (ADN et ARN), les solvants et les phospholipides se basent sur des concepts et des descriptions de la physique moléculaire très semblables au premier modèle – le *Consistent Force Field* – développé par Shneior Lifson, Arieh Warshel et Michael Levitt en 1967–1968 [268, 276]. Leur but est de décrire le plus simplement et le plus adéquatement possible les interactions liées et non-liées présentées à la Section 2.4.

Les trois principes de bases d'un modèle MM sont la transférabilité des paramètres, l'additivité des interactions et l'hypothèse thermodynamique. La transférabilité des paramètres permet de définir un nombre restreint de types d'atomes qui sont valides dans des contextes différents (e.g. un type d'atome peut être utilisé dans deux acides aminés différents). L'additivité des interactions permet de calculer indépendamment chaque interaction liée et non-liée, puis d'en faire l'addition pour obtenir l'énergie et la force totale sur chaque atome. L'hypothèse thermodynamique dit que l'état observé expérimentalement (e.g. la structure d'une protéine) est l'état de

[4] Le premier modèle gros-grain a été développé par Michael Levitt et Arieh Warshel en 1975 [270]. Dans celui-ci, chaque acide aminé est approximé par deux pseudo-atomes : un pour la chaîne latérale et un pour la partie du squelette de la protéine. Ce modèle a permis la première simulation du repliement d'une protéine.

plus basse énergie libre.

Durant les trente dernières années, plusieurs champs de force tout-atomes ont été développés pour les protéines dont AMBER [78], CHARMM [67, 69], OPLS [218, 397] et GROMOS [357][5] qui sont les plus communément utilisés de nos jours. Plusieurs champs de force gros-grains – composés de pseudo-atomes – ont aussi été développés et sont couramment utiliser dont OPEP [444], Martini [307], PaLaCe [369], UNRES [287] et PRIMO [223].

Puisqu'il s'agit tous de modèles empiriques, ils sont une approximation des phénomènes quantiques réellement présents à l'échelle nanométrique. Ils ne sont donc pas parfaits et chacun d'entre eux subit des améliorations constantes.[6] L'avènement d'ordinateurs toujours plus puissants permet, d'une part, de tester plus en profondeur des systèmes moléculaires variés sur des échelles de temps et d'espace plus importantes. De nouvelles techniques de paramétrisation sont, d'autre part, constamment développées pour corriger les champs de force empiriques et repousser leurs limites.

Dans cette section, le développement des champs de force AMBER et OPLS (tout-atome) ainsi que OPEP (gros-grain) est spécifiquement présenté. Les représentations des molécules d'eau et des phospholipides utilisés de concert avec AMBER dans les prochains chapitres sont aussi décrites.

3.1.1 Le champ de force tout-atome AMBER

Le développement du champ de force tout-atome AMBER a débuté en 1983 sous la supervision de Peter A. Kollman [499, 500]. De nos jours, une communauté de groupes de recherche améliore constamment AMBER ainsi que la suite d'outils de simulations et d'analyses qui portent le même nom [78]. Un sommaire du développement du champ de force AMBER spécifiquement pour les protéines[7] est présenté au Tableau 3.I. Dans cette thèse, le champ de force AMBER est utilisé aux Chapitres 9, 11 et 12.

La philosophie des champs de force AMBER est, depuis le début, d'avoir un nombre minimal de paramètres (types d'atomes et de paramètres pour les angles dièdres), une fonctionnelle

[5] GROMOS est un potentiel atome-uni pour lequel les hydrogènes des carbones aliphatiques sont intégrés à ces derniers afin d'accélérer les calculs en réduisant légèrement le nombre de particules.

[6] Les simulations utilisant la mécanique quantique font aussi intervenir des approximations bien qu'elles se voient attribuer le terme d'*ab initio*.

[7] Bien que ceci n'est pas traité dans la présente thèse, AMBER s'applique aussi très bien aux acides nucléiques (ADN et ARN).

Tableau 3.I – Sommaire du développement du modèle AMBER pour les protéines

Nom	Référence	Année	Modifications principales
ff83	[499]	1984	Naissance du champ de force AMBER (atome-uni).
ff85	[500]	1986	Version tout-atome de ff83.
ff94	[103]	1995	Reformulation complète de ff85.
ff96	[249]	1997	Même formulation que ff94, mais ajustement des paramètres pour les angles dièdres ϕ et ψ.
ff99	[490]	1999	Même formulation que ff94, mais ajustement des paramètres des angles dièdres dont ceux de ϕ et ψ.
ff94gs	[167]	2002	Même formulation que ff94, mais sans contribution énergétique pour les angles dièdres ϕ et ψ.
ff03	[141]	2003	Nouvelle paramétrisation de ff94 à partir d'un nouveau protocole de détermination des charges partielles.
ff99ϕ	[436, 437]	2005	Même paramètres que ff99, mais réutilisation des paramètres de l'angle dièdre ϕ de ff94 dans ff99.
ff99sb	[190]	2006	Même formulation que ff99, mais correction des paramètres des angles dièdres ϕ, ψ, ϕ' (C–N–C^α–C^β) et ψ' (C^β–C^α–C–N).
ff99sb*/ff03*	[46]	2009	Même formulation que ff99sb/ff03, mais correction des paramètres de l'angle dièdre ψ, sauf pour la glycine et la proline, avec le modèle d'eau TIP3P.
ff03w	[47]	2010	Même formulation que ff03, mais correction des paramètres de l'angle dièdre ψ, sauf pour la glycine et la proline, avec le modèle d'eau TIP4P/2005.
ff99sb-ildn	[281]	2010	Même formulation que ff99sb, mais correction des paramètres des angles dièdres χ_1 pour l'isoleucine et la leucine, et χ_1 et χ_2 pour l'aspartate et l'asparagine.
ff99sb-nmr	[272]	2010	Même formulation que ff99sb, mais ajustement des paramètres des angles dièdres ϕ et ψ.
ff99sb_$\phi\psi$	[273]	2011	Même formulation que ff99sb, mais ajout d'une énergie de couplage entre les angles dièdres ϕ et ψ.
ff99sb_ϕ'	[341]	2011	Même formulation que ff99sb, mais correction des paramètres de l'angle dièdre ϕ'.
ff99sb*-ildn	[382]	2011	Combinaison de ff99sb* et ff99sb-ILDN.
ff14sb	[78]	2014	Inspiré de ff99sb*-ildn et ff99sb-nmr. À partir de ff99sb, ajustements des paramètres des angles dièdres ϕ et ψ, ainsi que χ_1 et χ_2 des chaînes latérales de tous les acides aminés (sauf glycine et proline).

simple seulement composée d'interactions à deux corps, ainsi que des charges partielles non-polarisables centrées sur les atomes et déterminées par calculs *ab initio*.[8] L'énergie des interac-

[8]La signification du terme calculs *ab initio* est : calculs fondamentaux par mécanique quantique. Ces calculs bien que caractérisés d'*ab initio* font aussi intervenir des approximations.

tions atomiques dans AMBER est décrite par l'équation 3.1 et la force totale sur chaque atome est obtenue par différentiation de la fonction d'énergie. La forme de chacune de ces interactions est présentée à la Figure 3.2. Les trois premiers termes sont les interactions liées : liaisons covalentes, angles de valence et angles dièdres. Le dernier terme représente les interactions non-liées : van der Waals et électrostatiques. Les liaisons covalentes et les angles de valence sont décrits par des potentiels harmoniques centrés à une valeur d'équilibre (r_{eq} et θ_{eq}, respectivement) et caractérisés par une constante de rappel (k_r et k_θ, respectivement). Les angles dièdres sont décrits par une série de Fourier composée, au plus, des trois premiers termes. Les interactions de van der Waals[9] sont décrites par le potentiel de Lennard-Jones caractérisé par deux termes (A_{ij} et B_{ij}) qui contrôlent la distance idéale entre les deux atomes ainsi que l'énergie qui y est associée.[10] Les interactions électrostatiques sont décrites par le potentiel de Coulomb et elles sont contrôlées par les charges partielles (q_i et q_j) ainsi que par la constante de permittivité.[11] La forme du potentiel ne change pas à travers les diverses améliorations apportées à AMBER, mais les paramètres ainsi que les stratégies de paramétrisation évoluent.

$$\begin{aligned} \mathrm{E}_{\text{total}} &= \sum_{\text{liens}} k_r \left(r - r_{\text{eq}}\right)^2 + \sum_{\text{angles}} k_\theta \left(\theta - \theta_{\text{eq}}\right)^2 + \\ &\sum_{\text{dièdres}} \frac{V_n}{2} \left[1 + \cos\left(n\phi - \delta_n\right)\right] + \sum_{i<j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon r_{ij}}\right] \end{aligned} \tag{3.1}$$

Dans les deux premières versions de AMBER (**ff83** [499] et **ff85** [500]), les paramètres sont obtenus comme suit : (i) r_{eq} et θ_{eq} sont principalement mesurés à partir de structures obtenues grâce à la diffractométrie de rayons-X (DRX), (ii) k_r, k_θ et V_n sont ajustés pour que les modes normaux de plusieurs composés correspondent aux fréquences mesurées expérimentalement,

[9] Celles-ci se produisent seulement entre les atomes espacés par au moins 3 liaisons covalentes, sinon ce sont les interactions liées qui décrivent leurs interactions. À trois liaisons covalentes de distance, les interactions de van der Waals sont présentes, mais un facteur d'échelle de 1/2 lors est appliqué dans AMBER pour réduire la trop grande répulsion du terme répulsif des interactions de van der Waals.

[10] Le premier terme du potentiel de Lennard-Jones est le meilleur compromis entre rapidité de calcul et représentativité de la répulsion électronique entre deux atomes trop proches. Le potentiel de Buckingham est une autre forme couramment utilisée pour modéliser les interactions de van der Waals. Le premier terme de ce potentiel est un exponentiel décroissant qui modélise mieux la répulsion électronique au prix de calculs informatiques plus demandants dans un ratio très approximatif de 3 pour 2. Dans les deux formulations, le deuxième terme a la même forme et représente la force de dispersion de London (interaction dipôle induit–dipôle).

[11] La constante de permittivité est égale à 1 dans une simulation avec représentation explicite des molécules d'eau. Elle dépend de la distance lorsqu'une représentation implicite du solvant est utilisée.

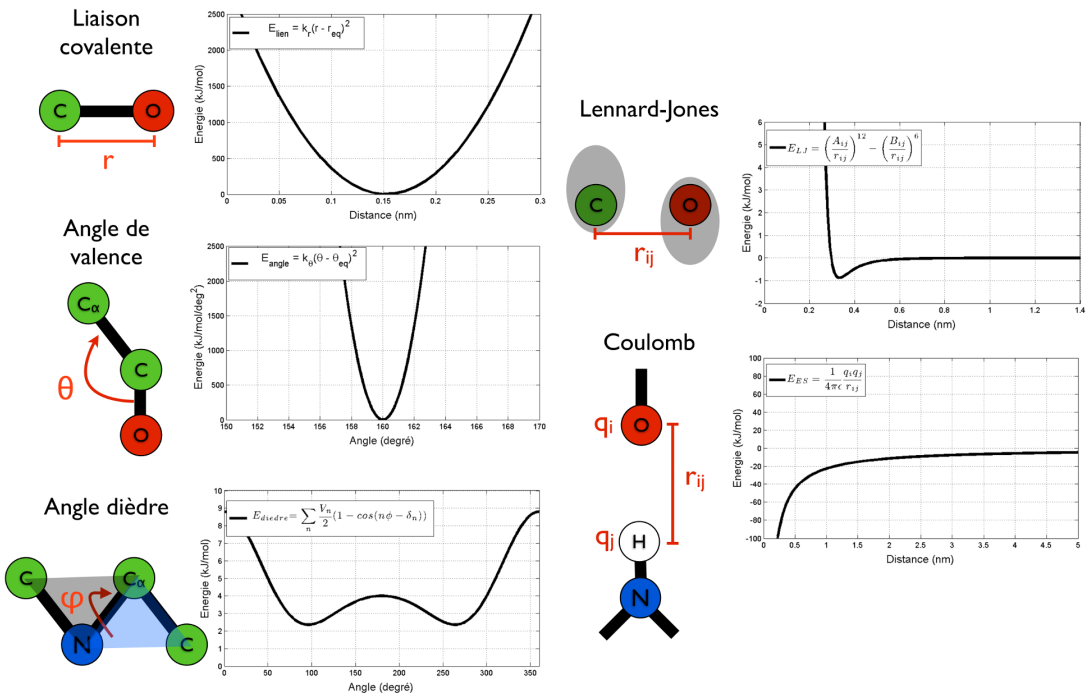


Figure 3.2 – Graphique des fonctions d'énergie des types d'interactions dans le modèle empirique tout-atome AMBER.

(iii) V_n est aussi ajusté pour que les différences d'énergie entre les stéréoisomères correspondent avec l'expérience, (iv) A_i et B_i sont extraits par l'analyse de l'empaquetage de cristaux,[12] et (v) les charges partielles sont obtenues par calculs *ab initio* avec la base STO-3G[13]. Dans ces versions, les ponts hydrogènes sont aussi explicitement représentés dans le potentiel afin d'ajuster plus précisément leur longueur.[14] Ces deux potentiels ont été calibrés et testés pour des simulations avec une représentation implicite du solvant ($\varepsilon = r_{ij}$).

La version **ff94** est une re-optimisation complète du potentiel tout en gardant la même philosophie [103]. Contrairement aux versions précédentes, elle est spécifiquement calibrée pour être compatible avec des simulations en solvant explicite – présence explicite des molécules d'eau – ce qui devenait peu à peu la norme à cette époque. Les cinq modifications fondamentales par rapport aux versions précédentes sont : (i) les charges partielles sont déterminées par calculs

[12] Les paramètres utilisés correspondent principalement à des mesures faites par d'autres groupes de recherche et ils ont été ajustés empiriquement lorsque nécessaire.

[13] Les deux limitations principales des calculs *ab initio* sont que les résultats varient selon la base et les conformations utilisées. De plus, à cette époque, les calculs étaient faits sur de petits fragments, puis extrapolés à la chaîne latérale et au squelette de chaque acide aminé.

[14] Le terme pour les ponts-H : $\sum_{\text{ponts-H}} \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}$.

ab initio en utilisant la base 6-31G* avec un protocole de restrictions,[15] (ii) les paramètres de van der Waals sont calculés à partir de simulations Monte Carlo de liquides organiques pour reproduire leurs valeurs expérimentales de densité et d'enthalpie de vaporisation,[16] (iii) les paramètres des angles dièdres sont déterminés, lorsque nécessaire, par calculs *ab initio* sur de petits composés et généralisés aux acides aminés de la protéine, (iv) les paramètres des angles dièdres ϕ et ψ sont déterminés par calculs *ab initio*,[17] et (v) le terme pour les ponts-H n'est plus nécessaire grâce aux meilleurs paramètres de van der Waals et charges partielles. Le protocole à suivre pour les optimisations subséquentes ainsi que pour la paramétrisation de nouveaux composés selon la philosophie AMBER est aussi présenté [103]. Il repose sur quatre étapes : (i) déterminer les charges partielles par calculs *ab initio* en utilisant la base 6-31G* et le protocole de restriction,[18] (ii) déterminer les paramètres van der Waals à partir des simulations Monte Carlo déjà faites sur de nombreux liquides, (iii) déterminer les paramètres des angles dièdres par calculs *ab initio*,[19] et (iv) ajuster les paramètres des liaisons covalentes et des angles de valence pour avoir la meilleure correspondance possible entre les calculs de modes normaux et les fréquences mesurées expérimentalement. En somme, les interactions non-liées – van der Waals et électrostatiques – ainsi que les paramètres des angles dièdres – particulièrement pour ϕ et ψ – sont considérés prioritaires afin d'obtenir un bon modèle empirique pour simuler la structure des protéines.

Depuis la version ff94, les améliorations suivantes ont principalement tenté de corriger le déséquilibre observé entre les diverses régions du diagramme de Ramachandran (ϕ,ψ) dans des simulations sur de petits peptides. La conséquence directe de ce déséquilibre est la déstabilisation de certaines structures secondaires par rapport à d'autres. Il va sans dire que cette situation

[15] Le choix de la base 6-31G* est justifiée par le fait qu'elle surestime uniformément la polarité moléculaire ce qui imite en quelque sorte l'environnement polaire dans lequel la protéine se trouve, contrairement à la phase gazeuse d'un calcul *ab initio*. Ceci permet, en quelque sorte, de balancer correctement les interactions protéine–protéine et solvant–protéine lorsque la protéine est simulée en présence explicite de molécules d'eau. Le protocole de restrictions est une amélioration cruciale, car il permet de corriger la présence de charges partielles trop importantes sur les atomes enfouis comme les carbones sp^3. De plus, avec ce protocole, les résultats dépendent moins de la configuration utilisée lors des calculs *ab initio*.

[16] Ceci a été inspiré des travaux sur le champ de force OPLS et s'inscrit dans la tendance de faire des simulations en présence explicite du solvant.

[17] Ces paramètres n'étaient pas présents dans les versions précédentes, car ils ne reposent sur aucune justification physique. Or, ils sont néanmoins nécessaires afin de calibrer le paysage énergétique de ϕ et ψ selon les charges partielles obtenues par calculs *ab initio*.

[18] Les charges sont déterminées à partir de la moyenne des résultats obtenus sur un ensemble de conformations de basses énergies.

[19] Grâce à l'approche minimaliste de AMBER, plusieurs de ces paramètres sont déjà disponibles.

est très problématique lorsque la stabilité et le repliement des structures tridimensionnelles des protéines sont étudiés. La première tentative (**ff96**) consistait à optimiser les paramètres des angles dièdres ϕ et ψ pour que les énergies relatives entre différentes conformations du di- et du tetra-peptide alanine[20] correspondent mieux avec les calculs *ab initio* [249]. Or, d'autres études ont démontré, par la suite, que ce potentiel surestime la stabilité des régions étendues (feuillets-β). Une deuxième tentative (**ff99**) plus sérieuse a ensuite été entreprise : optimisation des paramètres de tous les angles dièdres généraux sur un ensemble de 82 composés chimiques ainsi que des angles dièdres ϕ et ψ sur 6 conformations du di-peptide alanine et 11 conformations du tetra-peptide alanine pour lesquels plusieurs calculs *ab initio* précis sont disponibles [490]. La procédure d'optimisation – appelée Parmscan – trouve la meilleure série de Fourier en changeant systématiquement les valeurs de V_n afin de minimiser les différences d'énergies par rapport au jeu d'apprentissage. Or, d'autres études ont démontré par la suite que ce potentiel surestime la stabilité des régions hélicoïdales. Quelques tentatives simples ont été effectuées (**ff94gs** [167] et **ff99**ϕ [436, 437]), mais toutes les approches tentées possèdent un problème inhérent : les optimisations faites pour corriger ce déséquilibre des structures secondaires doivent prendre en compte les couplages entre ϕ (C–N–C^α–C) et ϕ' (C–N–C^α–C^β) autour du lien N–C^α ainsi qu'entre ψ (N–C^α–C–N) et ψ' (C^β–C^α–C–N) autour du lien C^α–C, car les angles dièdres ϕ' et ψ' sont absents de la glycine ce qui la distingue des autres acides aminés. Pour ce faire, le modèle **ff99sb** optimise les paramètres des angles dièdres ϕ et ψ en minimisant la différence d'énergie entre le champ de force empirique et les calculs *ab initio* pour 28 conformations du tetra-peptide glycine (Ac–Gly_3–NME), et ceux de ϕ' et ψ' pour 51 conformations du tetra-peptide alanine (Ac–Ala_3–NME) [190].[21] Ces paramètres sont ensuite appliqués à tous les acides aminés. L'équilibre résultant entre les structures secondaires PPII, β, α_L et α_R est significativement meilleur.

La version du potentiel AMBER utilisée aux Chapitres 9, 11 et 12 est **ff99sb*-ildn** [382] qui se base directement sur ff99sb. Elle corrige, d'une part, une déstabilisation légère des hélices-α (**ff99sb*** [46]) et elle ajuste, d'autre part, les paramètres des angles dièdres χ_1 pour l'isoleucine et la leucine ainsi que χ_1 et χ_2 pour l'aspartate et l'asparagine (**ff99sb-ildn** [281]). La première modification est une légère correction des paramètres de l'angle dièdre ψ afin de bien représenter la population d'hélices-α observée par RMN pour un petit peptide principalement désordonné

[20] Les peptides correspondent à Ac–ALA_n–NME où Ac est CH_3CO et NME est $NH(CH_3)$. Le nombre d'alanine (n) est respectivement 1 et 3 dans le cas du di- et du tetra-peptide.

[21] La glycine est, en effet, le seul acide aminé à ne pas avoir de C^β, donc à ne pas avoir de ϕ' et ψ'. L'alanine, quant à lui, est l'acide aminé le plus simple à avoir un C^β.

– Ac-$(AAQAA)_3$-NH_2 – à 300K.[22] La deuxième modification permet de réduire l'écart entre les valeurs d'énergie *ab initio* et celles du potentiel pour les angles dièdres de ces quatre acides aminés.[23]

Il existe aussi d'autres versions de AMBER. La version **ff03** est une re-paramétrisation complète des charges partielles selon un nouveau protocole à partir de ff99. L'effet de la polarisation de la protéine due au solvant est directement pris en compte lors des calculs *ab initio* des charges partielles en utilisant un continuum diélectrique qui imite l'effet du solvant sur la protéine.[24] Les paramètres des angles dièdres ϕ et ψ sont, pour leur part, re-optimisés selon les nouvelles charges partielles. Il existe aussi une version **ff03*** qui corrige une sur-stabilisation des hélices-α de ff03 [46]. La version **ff03w** est calibrée pour des simulations avec le modèle TIP4P/2005 pour les molécules d'eau[25] et elle provient de la même méthodologie que pour obtenir ff03* à partir de ff03 [47]. Le modèle résultant est caractérisé par une coopérativité accrue du repliement du peptide Ac-$(AAQAA)_3$-NH_2 en meilleure accord avec l'expérience que pour ff03 et ff03*. La version **ff99sb_ϕ'** introduit une légère modification à l'angle dièdre ϕ' (C–N–C_α–C_β) de ff99sb afin d'améliorer la correspondance entre les valeurs de couplage-J expérimentales et *in silico* sur Ala_3, Gly_3 et Val_3 à pH 2 en augmentant l'échantillonnage de configurations PPII [341].[26] La version **ff99sb-nmr** est dérivée à partir d'un nouveau protocole itératif et automatique qui ajuste les paramètres des angles dièdres ϕ et ψ de ff99sb afin d'augmenter la correspondance entre les valeurs expérimentales et *in silico* de déplacements chimiques pour un ensemble de protéines [272].[27] Le même protocole a aussi été appliqué pour optimiser une version modifiée de ff99sb qui inclut un nouveau terme d'énergie introduisant des corrélations entre les angles

[22] Les paramètres de l'angle dièdre ψ sont modifiés de la même façon pour tous les acides aminées. La correction n'est pas appliquée à la glycine et à la proline.

[23] Les auteurs ont observé que l'échantillonnage de l'angle dièdre χ_1 pour l'isoleucine, la leucine, l'aspartate et l'asparagine est le plus problématique en comparant les résultats de dynamiques moléculaires sur un petit peptide hélicoïdale – $(Ala)_4X(Ala)_4$ où X est n'importe quel acide aminé, sauf la glycine, l'alanine et la proline – et une analyse statistique des hélices dans la PDB [43]. Les calculs *ab initio* ont été faits avec la base MP2-aug-cc-pTZ et ils couvrent tout le spectre de valeurs pour χ_1 et χ_2.

[24] Au lieu de l'approche indirecte utilisée lors de la paramétrisation de ff99 : emploie de la base (6-31G*) qui surestime uniformément la polarité moléculaire afin d'imiter, en quelque sorte, l'environnement polaire dans lequel la protéine se trouve.

[25] Ce modèle reproduit mieux les propriétés de l'eau que TIP3P – un modèle plus simpliste qui est néanmoins celui recommandé pour AMBER – sur une plage large de températures et pressions. Le modèle ff03*, quant à lui, avait été obtenu selon des simulations avec TIP3P.

[26] Le calibrage a été fait avec le modèle d'eau TIP4P-Ew.

[27] Le calibrage a été fait avec le modèle d'eau SPC/E.

dièdres ϕ et ψ (**ff99sb_$\phi\psi$**) [273].[28]

Inspirée par la nature des corrections de ff99sb-nmr [272], ff99sb* [46] et ff99sb-ildn [281], une nouvelle version officielle de AMBER (**ff14**) a été développée en 2014. Des ajustements ont été apportés à ff99sb afin d'obtenir une meilleure correspondance : (i) entre les valeurs expérimentales de couplage-J et les simulations de petits peptides d'alanine (Ala_3 et Ala_5), et (ii) entre les calculs *ab initio* et le potentiel pour les angles dièdres des chaînes latérales de tous les acides aminés,[29] sauf la proline [78]. AMBER14 est, entre autres, assez précis pour replier 16 protéines α, β and α/β de 10 à 92 acides aminés avec des temps de repliement de la microseconde à la seconde [344].[30]

D'autres études ont identifié des améliorations qui pourraient être considérées dans les futures versions de AMBER : (i) les calculs *ab initio* des paramètres des angles dièdres et des charges partielles pourraient être améliorés par l'utilisation de bases plus précises [171, 221], (ii) la nature polarisante du solvant pourrait être prise en compte directement lors des calculs *ab initio* par une approche systématique et non pas indirectement en utilisant une base qui surestime uniformément la polarité moléculaire comme pour ff94/ff99 [81], (iii) les angles dièdres pourraient être plus adéquatement échantillonnés lors de leur paramétrisation par calcul *ab initio* afin d'augmenter la qualité de la surface d'énergie potentielle empirique [70], (iv) les paramètres de van der Waals pour la protéine devraient être déterminés plus systématiquement et exhaustivement selon des protocoles de simulation modernes [85],[31] (v) la dépendance à la température de la population d'hélices-α est incorrecte, ce qui pourrait provenir en partie d'un manque de directionnalité et de coopérativité des ponts-H [46], (vi) les paramètres des angles dièdres ϕ et ψ pourraient être différents pour chaque acide aminé considérant l'inter-dépendance du squelette et de la chaîne latérale [212], (vii) le choix du modèle utilisé pour les molécules d'eau semblerait crucial : TIP4P-Ew et TIP4P/2005 performant mieux que le modèle typiquement recommandé

[28] Le calibrage a été fait avec le modèle d'eau TIP3P.

[29] Les paramètres ont été déterminés de façon à minimiser l'influence de la configuration du squelette de l'acide aminé sur le résultat de la paramétrisation.

[30] Ces simulations ont été faites en utilisant une représentation implicite du solvant.

[31] L'objectif de la paramétrisation reste le même que pour ff94 : maximiser l'accord entre les simulations et l'expérience sur les valeurs de densité et d'enthalpie de vaporisation de plusieurs liquides organiques. Par contre, il faut être plus systématique en employant les mêmes techniques de simulation qui sont utilisées actuellement pour simuler les protéines, car cela influence les valeurs de densité et d'enthalpie mesurées. Plus précisément, lors de la paramétrisation, le calcul des interactions électrostatiques longues portées doit être fait avec PME et des corrections doivent être appliquées sur les interactions van der Waals longues portées pour corriger l'emploi d'un rayon de troncation.

pour AMBER (TIP3P) [47, 341],[32] (viii) la balance des interactions solvant–protéine semblerait cruciale pour bien représenter la structure des protéines désordonnées, les affinités protéine–protéine et l'énergie de solvation de composés [48, 342], et (ix) la formation de ponts salins semblerait surestimée par rapport à l'expérience [121].

3.1.2 Le champ de force tout-atome OPLS-AA

Le développement du champ de force OPLS à partir de 1978 a culminé avec la présentation d'un modèle atome-uni OPLS/AMBER en 1990 par le groupe de William L. Jorgensen [218]. La philosophie principale de OPLS repose sur l'ajuster les paramètres de van der Waals et les charges partielles pour reproduire les mesures expérimentales de densité et d'enthalpie de vaporisation de plusieurs liquides organiques. À ceci s'ajoute l'approche minimaliste de AMBER en conservant la même forme de potentiel (équation 3.1 et Figure 3.2) tout en minimisant le nombre de types de paramètres. Le développement de OPLS s'est poursuivi selon la même philosophie avec la création d'une première version tout-atome en 1996 (OPLS-AA [217]) ainsi que de deux versions subséquentes en 2001 (OPLS-AA\L [220]) et en 2015 (OPLS-AA\M [397]). Dans cette thèse, la version OPLS-AA\L sera utilisée au Chapitre 10.

Les paramètres liés du premier champ de force OPLS proviennent directement de AMBER83 (atome-uni), tandis que les paramètres non-liés ont été déterminés systématiquement par une nouvelle méthode novatrice. Elle repose sur la détermination des paramètres de van der Waals et des charges partielles de composés organiques représentatifs de tous les groupes fonctionnels dans le squelette et la chaîne latérale des acides aminés. Pour ce faire, les paramètres optimaux sont déterminés afin que la densité et l'enthalpie de vaporisation de ces composés organiques en phase liquide soient reproduites par des simulations de Monte Carlo.[33] Ces paramètres sont ensuite directement transférés sur les protéines. Cette méthode a, entre autres, été reprise partiellement par AMBER pour le développement de sa deuxième version majeure (ff94) au niveau des interactions de van der Waals, lesquelles n'ont pas changées dans les versions

[32] Les modèles TIP4P-Ew et TIP4P/2005 sont, contrairement à TIP3P, compatibles avec les techniques de simulations modernes – calculs électrostatiques longues portées par PME et corrections pour les interactions van der Waals au-delà du rayon de troncation – et ils reproduisent mieux les propriétés de l'eau. Le couplage TIP4P-Ew avec ff99sb_ϕ' produit des valeurs de déplacements chimiques en meilleur accord avec l'expérience pour Ala_3 et Gly_3 [341]. Le couplage TIP4P/2005 avec ff03w produit une coopérativité accrue du repliement d'un petit peptide principalement désordonné et de protéines ordonnées en meilleur accord avec l'expérience [47].

[33] La déviation moyenne entre OPLS et l'expérience sur ces quantités est de moins de 3% pour les 36 composés organiques utilisés.

subséquentes. Pour les acides aminés chargés, les paramètres ont été déterminés en analysant la structure et l'énergie de complexes ion–molécule par calculs *ab initio* et d'ions hydratés par simulation Monte Carlo. OPLS est principalement compatible avec le modèle d'eau TIP4P, mais aussi avec TIP3P et SPC.

La première version tout-atome, appelée OPLS-AA, applique la même philosophie que la version atome-uni [217]. Les paramètres pour les liaisons covalentes et les angles de valence proviennent principalement de AMBER85. Les paramètres de van der Waals et les charges partielles sont re-déterminés en suivant le même protocole qu'utilisé pour OPLS. Par contre, les paramètres des angles dièdres, incluant ceux pour ϕ, ψ et χ_1, sont déterminés précisément de façon à reproduire les calculs *ab initio*[34] sur divers composés en utilisant les paramètres non-liés calculés.

Les deux versions subséquentes de OPLS reprennent les paramètres non-liés de OPLS-AA et mettent l'accent sur l'amélioration des paramètres des angles dièdres. Plus précisément, OPLS-AA\L peaufine les paramètres des angles dièdres ϕ, ψ, χ_1 et χ_2 grâce à l'utilisation d'une base plus précise[35] pour les calculs *ab initio* [220]. L'optimisation des paramètres des angles dièdres ϕ et ψ se concentre proche des minimums de la surface d'énergie du di-peptide alanine (Ac-Ala-NME). Les paramètres obtenus permettent à leur tour de bien décrire la surface d'énergie du tetra-peptide alanine (Ac-Ala_3-NME) renforçant leur validité. Ces paramètres sont ensuite transférés aux autres acides aminés. Un protocole similaire est appliqué aux angles dièdres χ_1 et χ_2 pour tous les autres acides aminés. Les seuls paramètres non-liés modifiés sont les groupes contenant un souffre (cystéine et méthionine), car ces groupes étaient trop polarisés dans OPLS-AA.

La toute dernière version, OPLS-AA\M raffine encore plus les paramètres des angles dièdres ϕ et ψ en échantillonnant exhaustivement les surfaces d'énergies des di-peptides alanine et glycine (Ac-Ala-NME et Ac-Gly-NME) avec une base encore plus précise pour les calculs *ab initio* et en minimisant l'erreur renormalisée par un poids de Boltzmann.[36] Six angles dièdres sont paramétrés simultanément pour ces deux systèmes : ϕ (C–N–C_α–C), ϕ' (C–N–C_α–C_β) et ϕ'' (H_α–C_α–C–N) qui tournent tous autour du lien N–C_α et ψ (N–C_α–C–N), ψ' (C_β–C_α–C–N)

[34]La base utilisée est RHF/6-31G* pour l'optimisation géométrique et pour les calculs finaux.

[35]La base utilisée est HF/6-31G* pour l'optimisation géométrique et LMP2/cc-pVTZ(-f) pour les calculs finaux.

[36]La base utilisée est wB97X-D/6-311++G(d,p) pour l'optimisation géométrique et B2PLYP-D3BJ/aug-cc-pVTZ pour les calculs finaux.

et ψ" (H_α–C_α–C–N) qui tournent tous autour du lien C_α–C.[37] Les paramètres optimisés produisent une surface d'énergie ϕ/ψ pour le tetra-peptide alanine (Ac-Ala_3-NME) beaucoup plus similaire aux calculs *ab initio* que les versions OPLS précédentes.[38] Ces paramètres sont ensuite transférés aux autres acides aminés améliorant significativement, par rapport à l'expérience, les valeurs de couplage-J échantillonnées par Ac-X-NME où X est n'importe quel acide aminé. Un protocole similaire est appliqué aux angles dièdres χ_1 et χ_2 où le nombre de paramètres différents est restreint le plus possible. Les populations de conformères des chaînes latérales obtenues sont significativement plus proches de l'expérience que pour les deux versions précédentes.

En somme, la dernière version de OPLS-AA semble assez prometteuse : il reste à confirmer ces résultats par d'autres groupes indépendants. L'aptitude de OPLS-AA\M à échantillonner le repliement des protéines et l'ensemble structurel de peptides désordonnés reste aussi à confirmer.

3.1.3 Les champs de force pour les molécules d'eau

Le développement des modèles empiriques pour les molécules d'eau s'est fait en parallèle à celui des champs de force pour les protéines [176]. Les modèles d'eau les plus couramment utilisés pour des simulations de protéines en solvant explicite sont historiquement TIP3P [215], TIP4P [216] et SPC/E [39] ainsi que, plus récemment, TIP4P/2005 [1] et TIP4P-Ew [189]. Tous ont été développés en gardant une approche minimaliste avec une structure atomique rigide (liaisons covalentes et angle de valence fixes) et un nombre restreint de paramètres tel qu'illustré à la Figure 3.3. Le but de ces modèles est d'inclure l'effet moyen de la polarisation dans un potentiel effectif d'interactions à deux corps. Cinq caractéristiques principales les distinguent : le nombre de sites d'interactions et leur position, les paramètres géométriques de la molécule d'eau, les paramètres de van der Waals, les charges partielles et le traitement des interactions longues portées lors de la paramétrisation. Les modèles présentés ci-dessus utilisent tous le potentiel de Lennard-Jones $\left(\frac{A}{r^{12}} - \frac{B}{r^6}\right)$ pour décrire les interactions de van der Waals. Dans cette thèse, le modèle d'eau TIP3P est utilisé aux Chapitres 9, 11 et 12 conjointement avec AMBER et le modèle d'eau TIP4P est utilisé au Chapitre 10 conjointement avec OPLS-AA\L.

Le modèle SPC/E [39, 41] développé par Herman J. C. Berendsen possède trois sites d'interactions : un site centré sur chacun des trois atomes. Les hydrogènes possèdent une charge

[37] Ce couplage a aussi été observé précédemment lors de la paramétrisation du champ de force ff99sb de AMBER.

[38] Les auteurs notent même une amélioration significative par rapport aux champs de force ff99, ff99sb, ff99sb-nmr de AMBER.

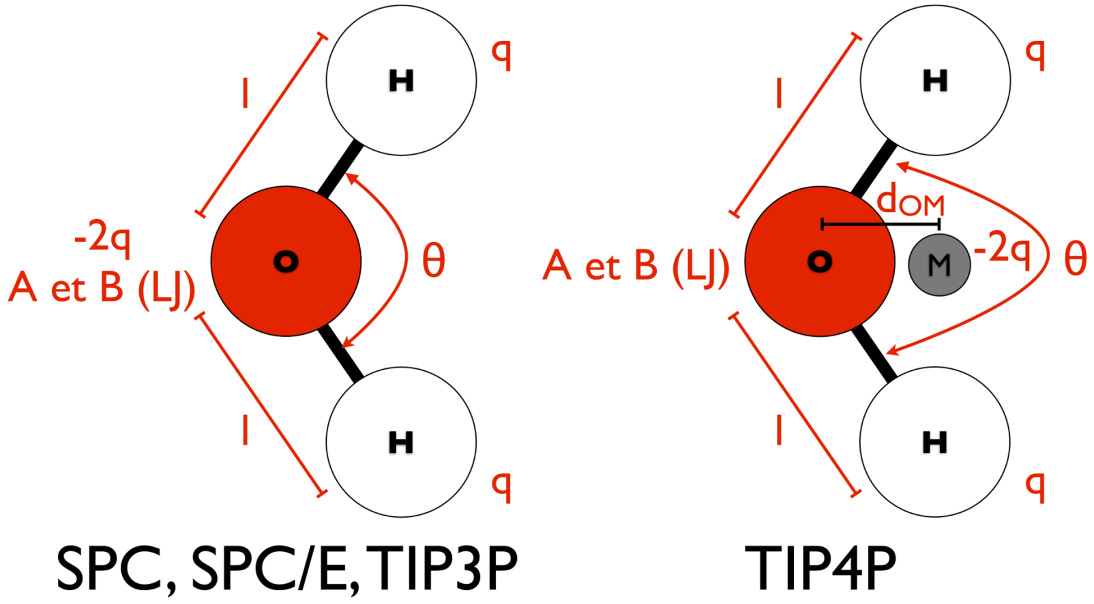


Figure 3.3 – Schématisation des modèles empiriques utilisés pour les molécules d'eau dans les simulations de protéines en présence explicite du solvant.

partielle q, mais pas de paramètres de van der Waals. L'oxygène possède des paramètres de van der Waals et une charge partielle $-2q$ afin que la charge totale de la molécule d'eau soit nulle. Dans le modèle, les valeurs de q et A sont optimisées, tandis la valeur *ab initio* $6,25456 \cdot 10^2$ kcal/mol·Å^6 est assignée à B. La géométrie de la molécule d'eau est fixée à $l = 1.0$ Å (liaisons H–O) et $\theta = 109,28°$ (angle HOH). Les valeurs idéales pour q et $A - q = 0,4238e$ et $A = 6,29358 \cdot 10^5$ kcal/mol·Å^{12} – sont déterminées de manière à ce que les mesures expérimentales de densité et d'énergie de vaporisation de l'eau soient reproduites. La distribution radiale de densité contient les deux pics caractéristiques de l'eau liquide et la constante de diffusion est en bon accord avec l'expérience à 300 K et 1 atm. Une caractéristique fondamentale de SPC/E est la prise en compte d'une correction de l'énergie interne due à un terme qui n'est pas présent dans les potentiels non-polarisants – le coût d'énergie pour polariser une molécule d'eau – lors des comparaisons avec les mesures d'énergie expérimentales.[39] Cette correction qui n'est pas présente dans la plupart des modèles d'eau pré-2000 est néanmoins nécessaire afin que les valeurs numériques sont cohérentes avec les mesures expérimentales. Les modèles plus récents comme TIP4P-Ew et TIP4P/2005 l'utilisent aussi.

[39] Cette correction distingue SPC/E de sa version antérieure, SPC.

Les modèles *transferable intermolecular potential* (TIP) TIP3P [215] et TIP4P [216], développés par le groupe de William L. Jorgensen, font aussi partie des plus populaires pour la simulation de protéines en solvant explicite. TIP3P est une version re-paramétrée de TIPS qui possède 3 sites d'interactions, un site sur chaque atome [213, 214]. Comme SPC et SPC/E, les hydrogènes possèdent une charge q et n'ont pas de paramètres de van der Waals, tandis que l'oxygène possède une charge $-2q$ et des paramètres de van der Waals. La géométrie de la molécule d'eau est caractérisée par $l = 0,9572$ Å et $\theta = 104,52°$, tandis que q, A et B sont optimisés de manière à ce que les propriétés structurelles et l'énergie de complexes en phase gazeuse et liquide soient respectivement en accord avec les calculs *ab initio* et l'expérience. Les paramètres pour TIP3P sont $q = 0,417e$, $A = 5,82000 \cdot 10^5$ kcal/mol·Å^{12} et $B = 5,95000 \cdot 10^2$ kcal/mol·Å^6. La distribution radiale obtenue par ce modèle ne possède pas, par contre, le deuxième pic caractéristique des molécules d'eau en phase liquide. Le modèle TIP4P résout cette problématique en ajoutant un quatrième site d'interaction (M) sur la bissectrice entre les deux liens O–H à une distance $d_{\mathrm{OM}} = 0,150$ Å de l'oxygène. Le site M possède la charge $-2q$ au lieu de l'oxygène qui garde seulement ses paramètres de van der Waals. Hormis ce quatrième site d'interaction, la géométrie de la molécule d'eau et le protocole d'optimisation des paramètres q, A et B sont les mêmes que pour TIP3P. Les paramètres pour TIP4P sont $q = 0,520e$, $A = 6,00000 \cdot 10^5$ kcal/mol·Å^{12} et $B = 6,10000 \cdot 10^2$ kcal/mol·Å^6. Les propriétés de l'eau liquide avec TIP4P autour de 300 K à 1 atm sont en meilleurs accord avec l'expérience. Entre autres, la distribution radiale de la densité possèdent les deux pics caractéristiques de l'eau liquide. Le désavantage de TIP4P par rapport à TIP3P provient de la présence du quatrième site, car ceci augmente le nombre d'interactions intermoléculaires à calculer durant une simulation. Par ailleurs, des modèles avec 5 sites existent comme TIP5P qui permettent de bien représenter les valeurs expérimentales de densité en fonction de la température [301].

Les modèles TIP3P et, plus particulièrement, SCP/E et TIP4P reproduisent assez bien un ensemble de propriétés de l'eau [176]. Ils ont cependant été paramétrés grâce à des simulations de dynamique moléculaire ou de Monte Carlo avec un traitement sous-optimal des interactions longues portées. Jadis, il était commun d'utiliser une troncation abrupte des interactions longues portées, même celles électrostatiques, pour des distances interatomiques plus grandes que 7,5–9,0 Å [39, 215, 216]. Or, ceci introduit des erreurs dans les quantités simulées qui doivent donc être corrigées par une approximation de champ moyen pour les interactions Lennard-Jones

longues portées (corrige l'énergie potentielle et la pression) et par l'utilisation de la méthode de Ewald pour calculer les interactions électrostatiques longues portées (corrige l'énergie). L'énergie doit aussi être corrigée dû au traitement classique du système durant la simulation afin de prendre en compte : (i) l'énergie de vibration intra- et inter-moléculaire, et (ii) l'énergie de polarisation des molécules d'eau comme lors de la paramétrisation du modèle SPC/E [39]. Toutes ces corrections sont nécessaires afin de comparer convenablement les valeurs d'énergie et de pression simulées avec les mesures expérimentales. Par ailleurs, il est aussi préférable de paramétrer le modèle de manière à ce qu'il puisse reproduire les propriétés de l'eau en fonction de la température. Deux modèles récents – TIP4P-Ew [189] et TIP4P/2005 [1], basés directement sur TIP4P – sont des exemples notables à cet égard. TIP4P-Ew est paramétré afin de reproduire la densité et l'enthalpie de vaporisation en fonction de la température. Les paramètres optimaux sont : $d_{\mathrm{OM}} = 0,125$ Å, $q = 1,04844e$, $A = 6,56138 \cdot 10^5$ kcal/mol·Å^{12} et $B = 6,53564 \cdot 10^2$ kcal/mol·Å^6. TIP4P/2005 est paramétré afin de reproduire le pic de densité en fonction de la température, la densité à condition ambiante, l'enthalpie de vaporisation et la stabilité de plusieurs polymorphes de glace. Les paramètres optimaux sont : $d_{\mathrm{OM}} = 0,1546$ Å, $q = 0,5564e$, $A = 7,31390 \cdot 10^5$ kcal/mol·Å^{12} et $B = 7,36095 \cdot 10^2$ kcal/mol·Å^6. Des comparaisons poussées sur un ensemble de propriétés montrent que ces deux modèles se démarquent des autres, avec un certain avantage pour TIP4P/2005 [1, 189, 479].

Historiquement, les champs de force pour les protéines sont habituellement associés à un modèle d'eau en particulier selon le modèle préféré lors de l'ajustement et de la validation des paramètres. Le champ de force AMBER, par exemple, est associé à TIP3P, tandis que CHARMM est associé à une version modifiée de TIP3P dans laquelle les hydrogènes possèdent des paramètres de van der Waals. OPLS-AA est principalement utilisé avec TIP4P et GROMOS avec SPC/E.

L'importance du choix du modèle d'eau pour les simulations de protéines a été négligée pendant plusieurs années, car l'effort était principalement mis sur l'amélioration des champs de force pour les protéines. Depuis que ces derniers ont atteint une certaine maturité, l'effet de la modélisation des molécules d'eau – ou de n'importe quel solvant d'ailleurs – a maintenant un impact significatif sur les propriétés structurelles des protéines. Entre autres, les valeurs de couplage-J et de déplacement chimique échantillonnées dans les protéines structurées sont influencées par le modèle d'eau. Il semble que TIP4P-Ew donne une meilleure correspondance avec l'expérience [341]. De plus, les protéines désordonnées naturellement ou par dénaturation

adoptent des structures plus compactes que ne le montrent les mesures expérimentales [380]. Il semble que TIP4P/2005 et TIP4P-D diminuent cet effet [47, 379]. D'autre part, il se peut que la transférabilité des paramètres d'une molécule d'eau dans un champ de force pour les protéines ne soit pas tout à fait adéquate : les interactions eau–protéine pourraient devoir être re-calibrées par rapport aux interactions protéine–protéine et eau–eau [48, 342]. Il s'agit clairement de nouvelles avenues de recherche qui devront être poursuivies plus en profondeur avant de changer définitivement les choix historiques du couplage des champs de force pour les protéines et l'eau.

3.1.4 Le champ de force gros-grain OPEP

Les champs de force tout-atomes comme AMBER et OPLS-AA offrent un bon degré de précision, mais ils requièrent beaucoup de puissance informatique malgré leur simplicité. Afin de simuler des processus complexes qui se produisent sur de longues échelles de temps comme l'oligomérisation ainsi que le repliement et la fonction de larges protéines, il est nécessaire de simplifier ces modèles tout-atomes. Pour y parvenir, les champs de force gros-grains comme OPEP mettent en pratique le concept de pseudo-atome qui consiste à regrouper certains atomes [444]. L'avantage est double : (i) moins d'atomes exige moins d'interactions à calculer, donc une rapidité de calculs accrue et (ii) une structure simplifiée des interactions donnent un paysage énergétique plus simple à explorer[40]. La clé d'un potentiel gros-grain est de réussir à simplifier tout en gardant la physique essentielle aux phénomènes étudiés. Dans cette thèse, le champ de force gros-grain OPEP est utilisé aux Chapitres 5, 6, 13 et VI.

La philosophie de OPEP repose sur une structure atomique minimaliste des chaînes latérales afin de seulement imiter leur caractère non-polaire ou chargé, et une structure atomique détaillée du squelette de la protéine afin de décrire finement les structures secondaires [123–126, 159, 311]. En effet, ceci semble offrir un degré de détails suffisant pour décrire convenablement le repliement et l'auto-assemblage des protéines. Ainsi, chaque acide aminé dans OPEP est modélisé par un seul pseudo-atome pour la chaîne latérale – sauf la proline qui est atome-uni[41] – et par tous les atomes du squelette – sauf l'hydrogène sur le carbone-α – tel qu'illustré à la Figure 3.4. Chaque acide aminé a donc 6 pseudo-atomes : SC (*side chain* pour chaîne latérale),

[40] La fonction de potentiel décrite par l'équation 3.1 peut, en effet, être vue comme un paysage fait de vallées et de monts. Les méthodes présentées à la prochaine section (Section 3.2) permettent d'explorer ce paysage énergétique afin de décrire les événements clés dans le repliement, l'oligomérisation et la fonction des protéines. Le paysage énergétique d'un modèle gros-grain est plus simpliste et, donc, plus facile à explorer pour ces méthodes.

[41] La proline dans OPEP ne contient que les atomes de carbones C_β, C_γ et C_δ au niveau de la chaîne latérale.

N, H_N, C_α, C et O, sauf la glycine et la proline. La glycine n'a pas de pseudo-atome SC, rôle qui est joué par son C_α. La proline représente les carbones de la chaîne latérale (C_β, C_γ et C_δ) et n'a pas d'hydrogène fixé à l'azote du squelette. De plus, le premier acide aminé au N-terminal n'a pas de H_N, tandis que le dernier acide aminé au C-terminal a un O supplémentaire.

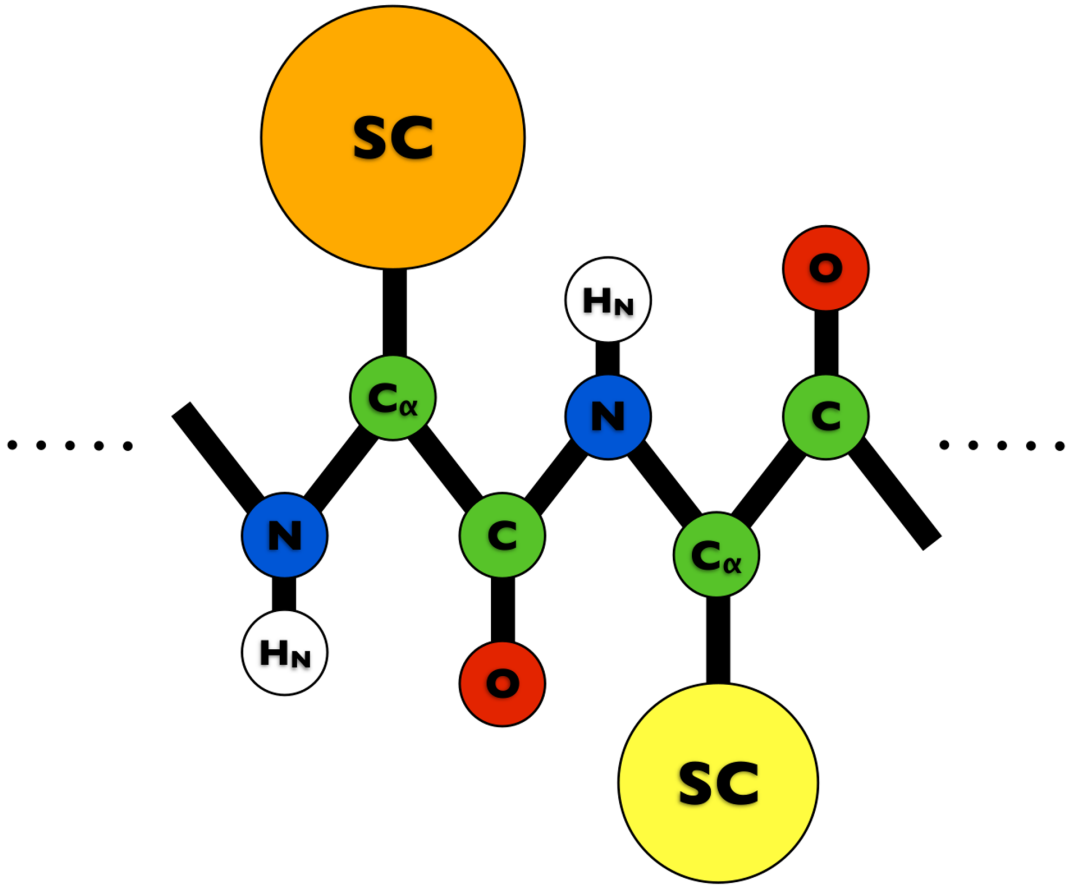


Figure 3.4 – Schématisation d'un acide aminé dans le champ de force gros-grain OPEP. Le pseudo-atome SC représente les atomes de la chaîne latérale.

Les premiers développements sur le champ de force OPEP ont débuté en 1997 par Philippe Derreumaux pour des simulations Monte Carlo de repliement [122–124]. La première version officielle, OPEPv1, a été optimisée sur un ensemble de leurres – la native (structure expérimentale) et des non-natives (structures incorrectement repliées) – afin de maximiser la différence entre l'énergie de la native et celles des non-natives [125–128]. Les concepts de ponts-H explicites et coopératifs ainsi que d'interactions SC-SC non-polaires étaient présents dans cette version. Elle a été utilisée, entre autres, dans les premières simulations de protéines avec ART, la

technique d'activation-relaxation initialement développée par Normand Mousseau pour échantillonner les événements d'activation dans les matériaux [334]. La deuxième version officielle, OPEPv2, reprend essentiellement la formulation de OPEPv1 avec des paramètres re-optimisés sur un ensemble de 6 protéines ayant de 10 à 38 acides aminés et testés sur un ensemble de 8 protéines possédant des motifs super-secondaires [129, 159].

Les fonctionnelles d'énergie et de force de OPEP ont passablement évolué dans les années suivantes grâce à l'utilisation de ART pour échantillonner l'espace énergétique de plusieurs protéines [130, 332, 494–496]. Ces travaux ont mené au développement d'une première version – OPEPv3 [311] – paramétrée grâce à un protocole d'optimisation strict. Cette version et celles subséquentes sont caractérisées (i) par l'absence de solvant explicite et implicite, car le repliement et l'auto-assemblage sont principalement déterminés par le caractère non-polaire des SC qui est pris directement en compte dans les interactions SC–SC, ainsi que (ii) par la présence d'interactions ponts-H directionnels et coopératifs. L'énergie des interactions dans OPEPv3 est ainsi décrite par l'équation 3.2.[42] La forme de chacune de ces interactions est présentée à la Figure 3.5. Chaque terme est multiplié par un poids (w) qui est optimisé lors de la calibration du potentiel. Les premiers termes sont les interactions liées – liaisons covalentes, angles de valence et angles dièdres – exprimées comme dans AMBER. Une quatrième interaction liée est ajoutée afin de restreindre harmoniquement les angles dièdres ϕ et ψ de tous les acides aminés dans les régions du graphique de Ramachandran qui leurs sont accessibles (Figure 3.5). La formulation des interactions non-liées est plus sophistiquée afin de palier l'absence de solvant et de charges partielles, et afin de décrire plus finement les ponts-H. Le premier terme non-lié (E_{vdw}) consiste en toutes les interactions van der Waals qui ne font pas intervenir deux pseudo-atomes SC ou deux C_α. La fonctionnelle utilisée est la même que pour AMBER où r^0_{ij} est la distance optimale entre les deux atomes et ε_{ij} est l'énergie associée à cette distance. Le deuxième terme non-lié ($E_{C\alpha–C\alpha}$) contrôle plus finement la distance idéale entre deux C_α espacés par au moins 5 acides aminés dans la séquence. Ceci est rendu nécessaire par la nature pseudo-atome de la chaîne latérale. Le troisième terme non-lié ($E_{SC–SC}$) décrit les interactions entre deux pseudo-atomes SC selon une formulation de style Lennard-Jones dans OPEPv3, laquelle est modifiée dans OPEPv4 et v5. Les interactions SC–SC sont soit répulsives–attractives (Lennard-Jones) ou

[42] Cette forme du potentiel correspond à la version 3.2 qui a servi de base pour les améliorations subséquentes des versions 4 et 5. La version 3.1 possède en plus un terme de propension intrinsèque à chaque acide aminé pour leur participation à la coopérativité des ponts-H. Un terme similaire existait aussi dans les versions précédentes OPEPv1 et OPEPv2.

seulement répulsives (*softcore*) selon les propriétés physico-chimiques des deux SC.[43] Le type de fonctionnelle utilisée est contrôlé par le signe de ε_{ij} et les fonctions Heaviside H(ε). Le quatrième terme non-lié (E_{HB}) décrit les interactions ponts-H entre les groupes donneurs (N–H_N) et accepteurs (O) du squelette de la protéine qui sont espacés par au moins 4 acides aminés dans la séquence.[44] La distance optimale des ponts-H est contrôlée par r^0_{HB} et l'énergie qui y est associée par ε_{ij} où cette valeur dépend du type de pont-H.[45] La dépendance en orientation des ponts-H est prise en compte par le $\cos^2 \alpha_{ijk}$ où α_{ijk} est l'angle entre les vecteurs formés par les liaisons covalentes H_N–N et $H_N \cdots O$ et où cette forme simple est choisie afin que l'énergie et la force tendent vers 0 lorsque l'angle tend vers 90°. L'énergie du pont-H est exactement zéro pour les angles plus petits que 90°. Le dernier terme non-lié (E_{HB2}) ajoute une coopérativité entre les ponts-H qui participent aux structures secondaires dominantes dans les protéines soient les hélices-α ainsi que les feuillets-β parallèles et anti-parallèles. Les topologies sont caractérisées par des paires de ponts-H (donneur : don, accepteur : acc) tel que (i) $\text{don}_1 = \text{acc}_1 + 4$, $\text{don}_2 = \text{acc}_2 + 4$ et $\text{acc}_2 = \text{acc}_1 + 1$ dans les hélices-α, (ii) $\text{don}_1 = \text{acc}_2$ et $\text{don}_2 = \text{acc}_1 + 2$ dans les feuillets-β parallèles, ainsi que (iii) $\text{don}_1 = \text{acc}_2$ et $\text{don}_2 = \text{acc}_1$ dans les feuillets-β anti-parallèles[46] (Figure 3.5).

[43] Les couples de chaînes latérales non-polaires ou de charges opposées sont attractifs, tandis que les autres couples sont pour la plupart répulsifs.

[44] Incidemment, il n'y a pas de ponts-H i/i+3 dans OPEP, donc pas d'hélice-3_{10}.

[45] Pour des considérations purement empiriques, trois catégories sont définies : i/i+4 (espacés par 4 acides aminés), i/i+5 (espacés par 5 acides aminés) et i/<i+5 (espacés par plus de 5 acides aminés) où les deux dernières sont associées au même poids *w*.

[46] Les paires de ponts-H satisfaisant $\text{don}_2 = \text{don}_1 + 2$ et $\text{acc}_2 = \text{acc}_1 + 2$ sont aussi considérées comme étant dans un feuillet-β anti-parallèle bien que ceci ne soit pas mentionné dans l'article sur OPEPv3.

$$
\begin{aligned}
\mathrm{E_{total}} &= \mathrm{E_{liées}} + \mathrm{E_{vdw}} + \mathrm{E_{C\alpha-C\alpha}} + \mathrm{E_{SC\text{-}SC}} + \mathrm{E_{HB}} + \mathrm{E_{HB2}} \qquad (3.2)\\
\mathrm{E_{liées}} &= w_\mathrm{b} \cdot \sum_{\text{liens}} k_r \left(r - r_\mathrm{eq}\right)^2 + w_\mathrm{a} \cdot \sum_{\text{angles}} k_\theta \left(\theta - \theta_\mathrm{eq}\right)^2 + \\
&\quad w_\mathrm{t} \cdot \sum_{\text{dièdres}} k_n \left[1 + \cos\left(n\phi - \delta_n\right)\right] + w_{\phi,\psi} \left[\sum_\phi k_\phi \left(\phi - \phi_0\right)^2 + \sum_\psi k_\psi \left(\psi - \psi_0\right)^2\right]\\
\mathrm{E_{vdw}} &= w_\mathrm{vdw} \sum_{\substack{i<j \\ \forall\{i,j\} \notin \\ \text{SC–SC} \cup \text{C}\alpha\text{–C}\alpha}} \varepsilon_{ij} \left[\left(\frac{r^0_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r^0_{ij}}{r_{ij}}\right)^{6}\right]\\
\mathrm{E_{C\alpha-C\alpha}} &= w_{\mathrm{C}\alpha-\mathrm{C}\alpha} \sum_{\substack{i+4<j \\ \forall\{i,j\} \in \text{C}\alpha\text{–C}\alpha}} \varepsilon_{\mathrm{C}\alpha-\mathrm{C}\alpha} \left[\left(\frac{r^0_{\mathrm{C}\alpha-\mathrm{C}\alpha}}{r_{ij}}\right)^{12} - 2\left(\frac{r^0_{\mathrm{C}\alpha-\mathrm{C}\alpha}}{r_{ij}}\right)^{6}\right]\\
\mathrm{E_{SC–SC}} &= \sum_{\substack{i<j \\ \forall\{i,j\} \in \\ \text{SC–SC}}} w^{\text{SC–SC}}_{ij}\, \varepsilon^{\text{SC–SC}}_{ij} \left[H(\varepsilon_{ij}) \left[\left(\frac{r^0_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r^0_{ij}}{r_{ij}}\right)^{6}\right] - H(-\varepsilon_{ij}) \left(\frac{r^0_{ij}}{r_{ij}}\right)^{6}\right]\\
\mathrm{E_{HB}} &= \sum_{\substack{\forall\{i,j\} \in \\ \text{N–H}_\text{N}\cdots\text{O}}} w^{\text{HB}}_{ij}\, \varepsilon^{\text{HB}}_{ij} \left[5\left(\frac{r^0_{\text{HB}}}{r_{ij}}\right)^{12} - 6\left(\frac{r^0_{\text{HB}}}{r_{ij}}\right)^{10}\right] \cos^2\left(\alpha_{\text{NHO}}\right) H\left(\alpha_{\text{NHO}} - 90^\circ\right)\\
\mathrm{E_{HB2}} &= \sum_{\substack{\forall\{i,j\} \in \text{N–H}_\text{N}\cdots\text{O} \\ \forall\{m,n\} \in \text{N–H}_\text{N}\cdots\text{O}}} w^{\text{HB2}}_{ij,mn}\, \varepsilon^{\text{HB2}}_{ij,mn} \exp\left(-\frac{\left(r_{ij} - r^0_{\text{HB}}\right)^2}{2}\right) \exp\left(-\frac{\left(r_{mn} - r^0_{\text{HB}}\right)^2}{2}\right)
\end{aligned}
$$

Dans **OPEPv3** [311], les paramètres liés ne faisant pas intervenir le pseudo-atome SC ont été essentiellement pris directement de AMBER91.[47] Les paramètres liés des chaînes latérales – la liaison covalente SC–C_α ainsi que les angles de valence N–C_α–SC et SC–C_α–C – ont été obtenus à partir d'une analyse statistique de 2 248 structures de la PDB avec moins de 30% d'identité de séquence. Le pseudo-atome SC, unique à chaque chaîne latérale, est placé au centre de masse des atomes réels en considérant la distribution des rotamères de la chaîne latérale. Les angles dièdres ϕ et ψ sont restreints harmoniquement dans la région définie par $\phi \in [-160^\circ, -60^\circ]$ et $\psi \in [-60^\circ, 160^\circ]$.[48] Les interactions non-liées $\mathrm{E_{vdw}}$ sont arbitrairement ajustés à un petit ε, car

[47] Le champ de force ff91 de AMBER est une version légèrement re-paramétrée des interactions non-liées de ff85 afin d'être compatible en présence explicite du solvant (TIP3P).

[48] La constante de rappel est de seulement 1,1 kcal/(mol·rad^2) pour tous les acides aminés, sauf pour la glycine et l'aspartate pour lesquels elle est 0,5 kcal/(mol·rad^2). Les acides aminés sont donc assez libres d'adopter n'importe

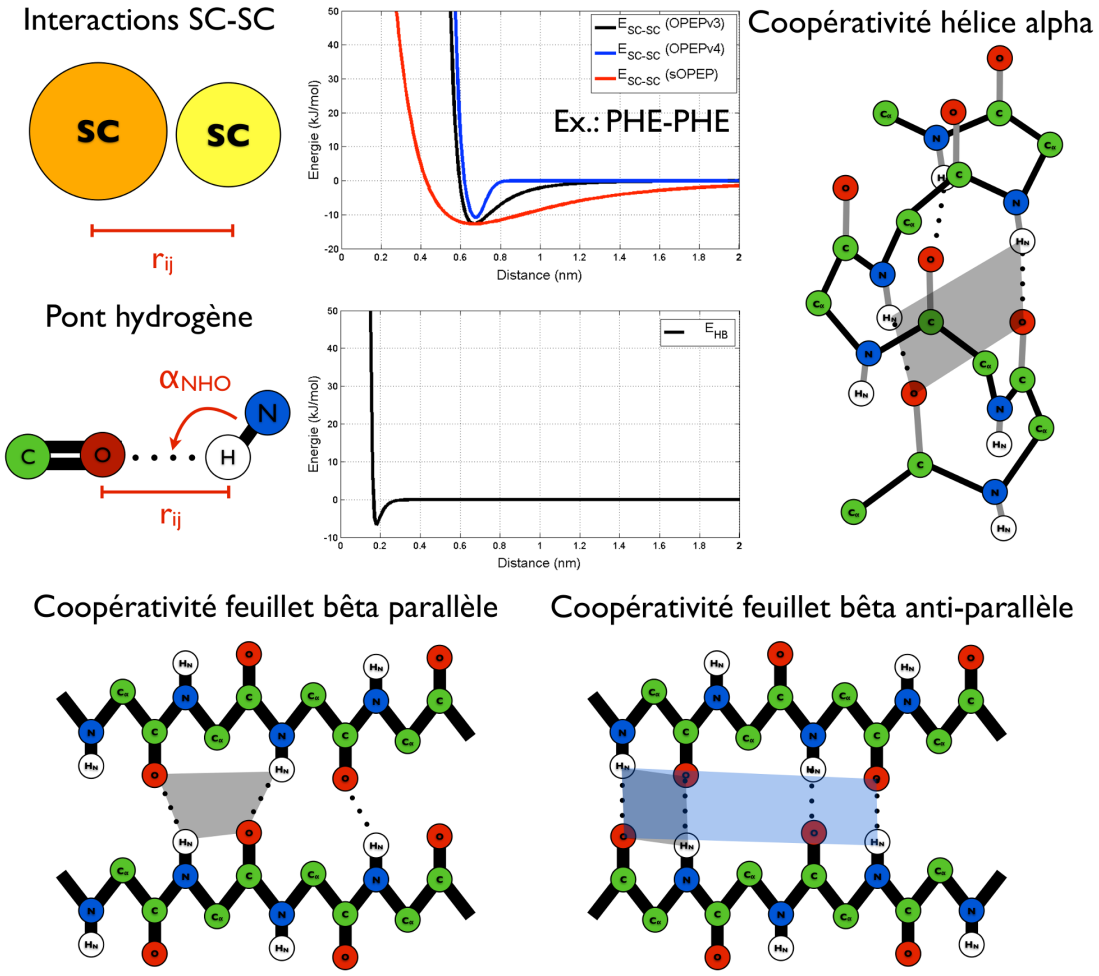


Figure 3.5 – Graphique de l'énergie des interactions chaîne latérale/chaîne latérale (SC–SC) et des ponts-H dans le modèle empirique gros-grain OPEP. Les topologies de ponts-H participant à la coopérativité des ponts-H sont aussi illustrées. Les autres interactions possèdent la même forme que dans le modèle AMBER (Figure 3.2).

ce terme sert seulement à bien représenter la taille finie des atomes dans OPEP. Les interactions C_α–C_α ont $r^0_{C\alpha-C\alpha}$ =5,6 Å et $\varepsilon_{C\alpha-C\alpha} = 0{,}199$ kcal/mol. La distance idéale C_α–C_α est évaluée à partir de l'analyse statique de la PDB précédemment énoncée. Dans OPEP, les interactions SC–SC, HB et HB2 sont dominantes, car elles sont les plus cruciales lors du repliement et de l'agrégation. L'énergie d'interaction de chaque couple SC–SC est contrôlée par $\varepsilon_{ij}^{\text{SC–SC}}$, lequel est obtenu à partir d'un potentiel statistique d'interaction [49].[49] La distance idéale de chaque couple SC–SC est évaluée à partir de l'analyse statistique de la PDB précédemment énoncée.

quelles valeurs de ϕ/ψ.

[49] Pour les interactions SC–SC attractive selon OPEP, la valeur $\varepsilon_{ij}^{\text{SC–SC}}$ est augmentée de 50%.

Les énergies de formation des ponts-H ($\varepsilon_{ij}^{\mathrm{HB}}$) sont 1,25 kcal/mol (j=i+4), 2,25 kcal/mol (j=i+5) et 2,0 kcal/mol (j>i+5) où les deux derniers types partagent le même poids w_{ij}^{HB}. La distance idéale d'un pont-H (r_{HB}^{0}) est 1,8 Å entre O et $\mathrm{H_N}$. Finalement, la coopérativité des ponts-H est caractérisée par un $\varepsilon_{ij}^{\mathrm{HB}}$ de 1,25 kcal/mol pour les hélices-α ainsi que de 2,0 kcal/mol pour les feuillets-β parallèles et anti-parallèles.

Le champ de force OPEPv3 est caractérisé par les 221 poids (w) montrés à l'équation 3.2 : 1 poids pour toutes liaisons covalentes (w_b), 1 poids pour tous les angles de valence (w_a), 1 poids pour tous les angles dièdres (w_t), 1 poids pour le potentiel ϕ/ψ ($w_{\phi,\psi}$), 1 poids pour les interactions van der Waals 1–4 et 1 poids pour toutes les autres ($w_{\mathrm{vdw}}^{1\text{–}4}$ et w_{vdw}),[50] 1 poids pour toutes les interactions C_α–C_α, 210 poids pour tous les couples possibles à partir des 20 types de chaînes latérales montrés à la Figure 2.3, 1 poids pour les ponts-H i/i+4 et 1 poids pour tous les autres ($w_{\mathrm{HB}}^{1\text{–}4}$ et w_{HB}), ainsi que 1 poids pour la coopérativité hélice-α et 1 poids pour la coopérativité feuillet-β parallèle ou anti-parallèle ($w_{\mathrm{HB2}}^{\alpha}$ et w_{HB2}^{β}). Ces poids sont ajustés de manière à ce que la structure expérimentale (native) ait la plus basse énergie lorsque comparée à des structures mal repliées, lesquelles sont séparées en deux catégories soient celles similaires à la native (native-proche) et les autres (non-native). Pour ce faire, un algorithme génétique est appliqué afin de maximiser le nombre de contraintes satisfaites sur un jeu d'apprentissage. Les poids évoluent, littéralement, à travers des mutations et des combinaisons de populations de poids candidats vers les solutions idéales. Le fonctionnement de cet algorithme génétique est expliqué plus en détail au Chapitre 13 lors de la présentation de aaOPEP.

Le jeu de protéines utilisé pour optimiser OPEPv3 contient 13 protéines avec des topologies α, β et α/β variées : 1ABZ (α), 1DV0 (α), 1E0M (β), 1ORC (α/β), 1PGB (α/β), 1QHK (α/β), 1SHG (β), 1SS1 (α), 1VII (α), 2CI2 (α/β), 2CRO (α), Betanova (β) et 1GPBF (β, acides aminés 41 à 56 de 1PGB). Ces protéines ne contiennent pas de ponts disulfures et d'acides aminés non-naturels. De 400 à 900 structures mal repliées sont générées pour chaque protéine : (i) par dynamique moléculaire à partir de la native, (ii) par repliement avec ART, (iii) par reconstruction glouton à partir de la séquence,[51] et (iv) par enfilage de la séquence sur d'autres

[50] Les interactions vdw 1–4 sont entre les paires d'atomes espacées par 3 liaisons covalentes, tandis qu'il n'y a tout simplement pas d'interaction vdw entre celles espacées par 1 ou 2 liaisons. Elles possèdent les mêmes paramètres que les autres interactions vdw, sauf qu'un facteur d'échelle de 1/80 leur est appliqué dans OPEP.

[51] Il s'agit de la procédure utilisée par PEPFOLD pour déterminer la structure des protéines à partir de la séquence [309, 310, 430, 461, 462]. La séquence est séparée en fragments de quatre acides aminés qui se recoupent. La structure de chaque fragment est prédite statistiquement selon une analyse de la PDB. Les fragments sont ensuite assemblés un à un et seulement les meilleures solutions sont gardées lors de l'ajout de chaque nouveau fragment

topologies. Un total de près de 8 000 leurres est ainsi utilisé lors de l'optimisation. La métrique du TM-score [509, 533] est utilisée pour classer les leurres en structures native-proches (TM $>$ 0.5) et non-natives.

Le pouvoir discriminant de OPEPv3 a été évalué et confirmé sur un ensemble de 16 protéines pour un total de plus de 20 000 leurres [311]. Son habilité à simuler le repliement et l'ensemble de conformations de protéines variées a été établi par plusieurs méthodes [444] dont ART [441], dynamique moléculaire d'échanges de répliques (REMD) [86], métadynamique (MetaD) [18] et prédiction *de novo* [310, 430]. OPEPv3 a permis de décrire l'agrégation de plusieurs protéines amyloïdes [131, 333, 497] telles que IAPP [256, 257], Aβ [88, 108, 109, 291, 314], α-synucléine [151], polyglutamine [255], NNQQ [292] et GNNQQNY [337, 338].

La version **OPEPv4** modifie la fonctionnelle des interactions SC–SC de OPEPv3 et ajoute 11 interactions SC–SC spéciales qui sont régulièrement présentes dans les hélices-α des protéines [89]. La nouvelle fonctionnelle est

$$\begin{aligned} \mathrm{E}_{\text{SC–SC}} \quad = \quad & \sum_{\substack{i<j \\ \forall\{i,j\} \in \text{SC–SC}}} w_{ij}^{\text{SC–SC}}\, \varepsilon_{ij}^{\text{SC–SC}} \left[H(\varepsilon_{ij}) \left(\frac{G(r_{ij}^0)}{r_{ij}} \right)^6 \mathrm{e}^{-2r_{ij}} + \right. \\ & \left. 0.6563701 \cdot \left(\tanh\left[2\left(r_{ij} - r_{ij}^0 - 0.5 \right) \right] - 1 \right) - H(-\varepsilon_{ij}) \left(\frac{r_{ij}^0}{r_{ij}} \right)^8 \right] \end{aligned} \tag{3.3}$$

où $G(r_{ij}^0)$ est une fonction qui permet de contrôler la distance entre les deux atomes pour une certaine énergie d'interaction,[52] et où tous les autres paramètres ont la même définition qu'auparavant. Cette fonctionnelle est inspirée des développements d'un modèle gros-grain OPEP pour les ARN et ADN [370] et elle permet d'obtenir un puits d'interaction plus serré lorsque comparé à la fonction de Lennard-Jones tel qu'illustré à la Figure 3.5.[53] De plus, pour les interactions SC–SC répulsives, cette nouvelle fonction est moins répulsive que la précédente. Les 11 nouvelles interactions SC–SC spéciales sont : Lys–Glu, Lys–Asp, Glu–Arg et Asp–Arg pour

(algorithme glouton).

[52] Plus précisément, $G(r_{ij}^0) = -0.7 \exp\left[2\left(\frac{r_{ij}^0 - 0.5}{5.0} \right) \right] \left(r_{ij}^0 - 0.5 \right)$ où la valeur de r_{ij}^0 est en Angströms. Par ailleurs, les fonctions $\mathrm{E}_{\text{SC–SC}}$ et $G(r_{ij}^0)$ de l'article ne correspondent pas à ce qui est écrit dans le code. Ici, les fonctions du code sont montrées.

[53] L'utilisation du tanh n'est pas vraiment idéale computationnellement, car il s'agit d'une fonction assez coûteuse à calculer. Une simple fonction switch aurait pu être utilisée pour obtenir le même effet.

i/i+3 et i/i+4 ainsi que Lys–Gln, Lys–Leu, Ala–Arg, Ala–Gln, Ala–Glu, Leu–Glu et Ile–Lys pour i/i+4. Un nouveau poids est associé à chacune de ces interactions pour un total de 11 poids supplémentaires. Les valeurs des poids finaux sont essentiellement les mêmes que pour OPEPv3 : seulement les 11 nouveaux poids et les 2 poids de la coopérativité sont optimisés par l'algorithme génétique. OPEPv4 conserve la stabilité structurelle de 17 protéines variées en dynamique moléculaire et replie des motifs variés en dynamique moléculaire d'échanges de répliques.

La version **OPEPv5** reprend exactement OPEPv4, sauf pour les interactions entre les chaînes latérales ayant des charges opposées : Arg–Asp, Arg–Glu, Lys–Asp et Lys–Glu [445]. Le potentiel de force moyenne entre ces paires est calculé par la méthode itérative d'inversion de Boltzmann à partir de simulations tout-atomes OPLS-AA\L en solvant explicite (TIP3P) à 300 K et 1 atm. Par la suite, le facteur d'échelle idéal à appliquer à ces nouvelles interactions a été déterminé grâce à plusieurs dynamiques moléculaires d'échanges de répliques. Ceci a permis d'ajuster leur échelle d'énergie par rapport aux autres interactions du potentiel qui n'ont pas été modifiées par rapport à OPEPv4.

À partir de OPEPv3, une autre famille de potentiels OPEP a été créée, celle du potentiel **sOPEP**. Ce modèle est utilisé dans la procédure PEPFOLD, qui reconstruit la structure tridimensionnelle d'une protéine seulement à partir de sa séquence [309, 310, 430, 461, 462], et dans la procédure PEP-SiteFinder, qui trouve les poses peptide–protéine idéales [407]. Dans sOPEP, les interactions SC–SC ne sont plus décrites par la fonction de Lennard-Jones, tandis que tous les autres termes sont exactement les mêmes que pour OPEPv3 tels que décrits par l'équation 3.2 [310].[54] La nouvelle fonction pour les interactions SC–SC sert à mieux représenter les distributions de distances observées expérimentalement afin de réduire le nombre de collisions interatomiques non-naturelles lors de la reconstruction tridimensionnelle faite par PEPFOLD. Elles sont maintenant décrites par

[54]Le potentiel sOPEP possède aussi, comme OPEPv3.1, un terme propension intrinsèque à chaque acide aminé pour leur participation à la coopérativité des ponts-H.

$$\mathrm{E}_{\text{SC–SC}} = \sum_{\substack{i<j \\ \forall\{i,j\} \in \\ \text{SC–SC}}} w_{ij}^{\text{SC–SC}}\, \varepsilon_{ij}^{\text{SC–SC}} \left[H(\varepsilon_{ij}) \left(\left(\frac{r_{ij}^0 - p_{ij}}{r_{ij} - p_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^0 - p_{ij}}{r_{ij} - p_{ij}} \right)^{6} \right) - H(-\varepsilon_{ij}) \left(\frac{2 \cdot R_{ij}^0 - r_{ij}^0}{r_{ij}} \right)^{6} \right] \tag{3.4}$$

où les valeurs de R_{ij} et de p_{ij} sont choisies à partir d'une analyse de la distribution des distances de tous les types de paires SC–SC (210 possibilités) pour 2 248 structures de la PDB ayant moins de 30% d'identité de séquence. Plus précisément, R_{ij} correspond à la distance représentant le quantile 0,1 de chacune de ces distributions et la valeur de p_{ij} est choisie pour que $E(R_{ij}) = 0$.[55] Le contrôle supplémentaire donné par R_{ij}^0 permet ainsi de réduire le nombre de collisions interatomiques non-naturelles lors de la reconstruction des modèles tridimensionnels à partir de la séquence par PEPFOLD [310].

Au Chapitre 13, un nouveau jeu d'apprentissage et des perspectives de développements futurs pour sOPEP sont présentés. Une extension tout-atome de la philosophie de OPEP est aussi développée pour, entre autres, peaufiner les prédictions *de novo* de PEPFOLD.

3.2 Les algorithmes d'échantillonnage

Le deuxième élément essentiel des méthodologies d'expériences *in silico* est un ensemble d'algorithmes qui explorent l'espace énergétique accessible à la protéine selon le champ de force utilisé [410]. Plusieurs logiciels versatiles contiennent une panoplie d'algorithmes et de champs de force tels que GROMACS [42, 185, 278, 387, 478], NAMD [377], DESMOND [63], OpenMM [145], AMBER [78] et CHARMM [67, 69] pour des calculs hautes performances sur des superordinateurs. D'autres logiciels se spécialisent sur un ensemble de méthodes en particulier et servent d'extension aux logiciels versatiles. Par exemple, le logiciel PLUMED se spécialise dans les méthodes de calculs d'énergie libre comme la métadynamique et il est compatible avec GRO-

[55] Cette condition sur l'énergie implique que $p_{ij} = \frac{r_{ij}^0 - \sqrt[6]{2} R_{ij}^0}{1 - \sqrt[6]{2}}$. Le quantile utilisé dans PEPFOLD pour spécifier R_{ij}^0 est le même que mentionné précédemment, bien que l'article sur sOPEP spécifie qu'il dépend du type d'interaction : 0,1 pour une interaction SC–SC attractive et 0,2 pour une interaction SC–SC répulsive. Le choix de réduire le quantile pour les interactions SC–SC répulsives permet d'éviter le rejet de structures tridimensionnelles acceptables.

MACS, NAMD, AMBER et LAMMPS [62, 467].

Dans cette section, les algorithmes utilisés dans les Chapitres 5, 6, 9, 10, 11 et 12 sont présentés. Certains permettent seulement une exploration locale de la surface d'énergie comme les algorithmes de minimisation et la dynamique moléculaire[56], tandis que d'autres permettent une exploration globale comme la dynamique moléculaire d'échanges de répliques et la métadynamique.

3.2.1 La minimisation

Les algorithmes de minimisation relaxent le système vers un minimum local d'énergie. Plusieurs méthodes existent, dont les algorithmes du gradient et du gradient conjugué ainsi que le *Fast inertial relaxation engine* qui seront utilisés dans les prochains chapitres.

L'**algorithme du gradient** (SD, *steepest descent*) et l'**algorithme du gradient conjugué** (CG, *conjugate gradient*) exécutent itérativement des déplacements vers le minimum local d'énergie selon une direction de descente re-déterminée à chaque pas. Ils consistent en 5 étapes successives à partir d'un arrangement initial de N atomes ($\vec{x}_0$, le vecteur des positions de dimension 3N) : (i) vérifier si la position actuelle des atomes ($\vec{x}_n$) minimise la fonction d'énergie et continuer seulement si ce n'est pas le cas, (ii) déterminer une direction de descente ($\vec{p}_n$) selon l'algorithme utilisé, (iii) déterminer le pas de la descente (λ_n) par recherche linéaire, (iv) calculer la nouvelle position $\vec{x}_{n+1} = \vec{x}_n + \lambda_n \vec{p}_n$, et (v) retourner à la première étape.

Le critère de convergence consiste principalement à demander que le gradient, lequel tend vers zéro à l'approche du minimum, soit plus petit qu'une certaine valeur. D'autres critères peuvent être ajoutés afin de s'assurer que la méthode de minimisation progresse suffisamment pour continuer. Définissant $\vec{g}_n = \nabla_{\vec{x}} E(\vec{x})\big|_{\vec{x}_n}$, la direction de descente est donnée par $\vec{p}_n = -\vec{g}_n$ pour SD et par $\vec{p}_n = -\vec{g}_n + \beta_n \vec{p}_{n-1}$ pour CG où $\beta_n = \vec{g}_n^{\mathsf{T}} (\vec{g}_n - \vec{g}_{n-1}) / \vec{g}_{n-1}^{\mathsf{T}} \vec{g}_{n-1}$ selon la méthode de Polak-Ribière[57]. L'algorithme SD procède donc toujours dans la direction opposée du gradient de l'énergie, c'est-à-dire dans la direction de la force, tandis que l'algorithme CG procède dans la direction perpendiculaire à toutes les directions de descentes précédentes. Le pas de la descente correspond à la valeur λ_n que la méthode de Newton-Raphson ou par la méthode de la

[56] L'étendue de l'échantillonnage de la dynamique moléculaire dépend ultimement du temps de simulation et/ou du nombre de simulations exécutées en parallèle.

[57] Initialement, $\vec{p}_0 = -\vec{g}_0$. Dans le logiciel GROMACS, la méthode Fletcher-Reeves avec $\beta_n = \vec{g}_n^{\mathrm{T}} \vec{g}_n / \vec{g}_{n-1}^{\mathrm{T}} \vec{g}_{n-1}$ est utilisée.

sécante détermine pour minimiser $E(\vec{x}_n + \lambda_n \vec{p}_n)$. Dans GROMACS, par contre, le pas de la descente pour l'algorithme SD est déterminé à partir de $\lambda_n = h_n / \max_i(\vec{F}(\vec{X}_n))$ où le module le plus grand de la force totale sur un atome est utilisé. La valeur de h_n est initialement choisie ($h_0 =$ 0,01 nm), puis elle augmente (diminue) par un facteur de 1,2 (0,2) lorsque l'énergie diminue (augmente).

L'algorithme SD est efficace loin du minimum, mais il y converge très lentement une fois proche de celui-ci. À l'opposé, CG converge plus rapidement, car les nouvelles directions de descentes sont perpendiculaires aux précédentes : la convergence est assurée, dans le cas d'une fonction quadratique convexe, en N pas où N est la dimension de l'espace de paramètres de la fonction à minimiser.[58] Dans le cas de l'énergie potentielle d'une protéine, le paysage énergétique est significativement plus complexe, ce qui nécessite de remettre $\beta_n = 0$ (i.e. enlever l'historique de descente par l'exécution d'un pas SD) au moins tous les N pas où N est la dimension de l'espace de paramètres de la fonction à minimiser. La méthode Polak-Ribière de CG nécessite aussi de remettre $\beta_n = 0$ lorsque la valeur déterminée est négative afin d'assurer la stabilité de la méthode.

L'algorithme ***Fast inertial relaxation engine*** (FIRE) fait partie de la catégorie des dynamiques moléculaires amorties [56]. Le principe de base de FIRE est d'utiliser l'inertie du système d'atomes en décrivant sa dynamique par : $\dot{\vec{v}}(t) = \vec{F}(t)/m - \gamma(t)v(t)\left[\hat{v}(t) - \hat{F}(t)\right]$ où $\vec{v}(t)$ est la vitesse, $\vec{F}(t) = -\nabla_{\vec{x}} E(\vec{x})$ est la force, m est la masse,[59] γ(t) est un paramètre qui contrôle la perturbation par rapport à la dynamique Newtonienne, v(t) est le module de la vitesse, $\hat{v}$ est le vecteur unitaire dans la direction de la vitesse et $\hat{F}$ est le vecteur unitaire dans la direction de la force. Tous les vecteurs contiennent 3N entrées où N est le nombre d'atomes.

Le paramètre $\gamma(t)$ est ajusté pour que la dynamique du système (vitesse) soit réorientée dans la direction de la force pour laquelle la descente est plus abrupte. Plus précisément, FIRE consiste en 4 étapes successives à partir d'un arrangement initial de N atomes ($\vec{x}_0$, le vecteur des positions de dimension 3N) ayant tous une vitesse nulle : (i) exécuter un pas de dynamique moléculaire standard pour déterminer la force, la vitesse et la position au temps suivant, (ii) vérifier si la position actuelle des atomes ($\vec{x}_n$) minimise la fonction d'énergie et continuer seulement si ce n'est pas le cas, (iii) calculer le signe de la puissance $P = \vec{F} \cdot \vec{v}$ qui indique l'orientation

[58] Sous réserve que la recherche linéaire – méthodes de Newton-Raphson ou de la sécante – converge assez proche du minimum dans chaque direction de descente.

[59] FIRE assume que chaque degré de liberté est comparable, donc la masse de chaque atome doit être la même lors de la minimisation.

relative de $\vec{F}$ par rapport à $\vec{v}$, (iv) échelonner la vitesse selon le terme γ(t) dans l'équation du mouvement par $\vec{v} \rightarrow (1-\alpha)\vec{v} + \alpha\hat{F}|\vec{v}|$ où $\alpha = \gamma\Delta t$,[60] (v) optimiser les paramètres α et Δt selon le signe de P, et (vi) retourner à la première étape.

L'avantage principal de FIRE est qu'il repose sur l'inertie et la dynamique des atomes du système, ce qui permet d'éviter de rester pris dans des minimums locaux peu profonds et de converger vers un minimum local plus pertinent. De plus, il converge plus rapidement que les méthodes traditionnelles telles que L-BFGS et CG pour des surfaces d'énergie assez complexes. La convergence, quant à elle, est établie selon des critères similaires à SD et CG tels que le gradient soit plus petit qu'une certaine valeur.

3.2.2 La dynamique moléculaire

La dynamique de la protéine peut aussi être analysée en intégrant les équations du mouvement. La grande majorité des simulations de protéines en présence explicite du solvant utilise la dynamique Newtonienne pour décrire le mouvement, la vitesse et l'accélération de chaque atome dans le système. En présence implicite du solvant, la dynamique de Langevin est habituellement privilégiée.

La **dynamique Newtonienne** d'un système de N atomes dans l'ensemble microcanonique (NVE) – nombre d'atomes, volume et énergie constants – repose sur l'équation du mouvement

$$\begin{aligned} \dot{\vec{X}}(t) &= \vec{V}(t) \\ \mathbf{M}\dot{\vec{V}}(t) &= -\nabla_{\vec{X}} E(\vec{X}) \end{aligned} \tag{3.5}$$

où $\mathbf{M}$ est une matrice 3N×3N ayant la masse de chaque atome sur la diagonale, $\ddot{\vec{X}}(t)$ est un vecteur 3N contenant le vecteur accélération de chaque atome, et $\vec{F}(\vec{X}(t)) = -\nabla_{\vec{X}} E(\vec{X})$ est un vecteur 3N contenant la force de chaque atome calculée à partir de l'énergie potentielle $E(\vec{X})$. Une fois la force calculée, l'équation du mouvement peut être résolue indépendamment pour chaque atome.[61] L'équation 3.5 est intégrée numériquement pour obtenir la vitesse $\dot{\vec{X}}_{t+\Delta t}$ et la

[60] Un P positif indique une direction favorable, donc α est diminué et Δt est augmenté afin de diminuer l'influence du terme γ sur la dynamique. Ces changements se produisent seulement si P est positif depuis un certain nombre de pas afin d'assurer la stabilité de la méthode. Un P négatif indique une direction non-favorable, donc α retourne à sa valeur initiale et Δt est diminué afin d'augmenter l'influence du terme γ sur la dynamique.

[61] Ceci n'est plus le cas lorsque la température ou la pression sont contrôlées ou lorsque des restreintes sont

position $\vec{X}_{t+\Delta t}$ après un petit pas de temps Δt à partir de la vitesse $\dot{\vec{X}}_t$ et de la position $\vec{X}_t$ au temps précédent ainsi que de l'accélération $\ddot{\vec{X}}_t$ qui est reliée à la force $\vec{F}(\vec{X}(t)) = -\nabla_{\vec{X}} E(\vec{X})\big|_{\vec{X}_t}$,

Il existe plusieurs méthodes d'intégration ayant différentes propriétés dont l'ordre de précision envers le pas d'intégration $\mathscr{O}(\Delta t^n)$. La plus populaire est sans contredit la famille d'intégrateurs de Verlet, car elle conserve l'énergie tout en ayant un ordre de précision raisonnable, $\mathscr{O}(\Delta t^2)$ [481]. Dans cette famille, l'algorithme Verlet vitesse [452] consiste en une séquence de 3 étapes par pas de temps Δt selon

$$\begin{aligned} \vec{V}_{t+\Delta t/2} &= \vec{V}_t + \frac{\Delta t}{2}\mathbf{M}^{-1}\vec{F}_t \\ \vec{X}_{t+\Delta t} &= \vec{X}_t + \Delta t \vec{V}_{t+\Delta t/2} \\ \vec{V}_{t+\Delta t} &= \vec{V}_{t+\Delta t/2} + \frac{\Delta t}{2}\mathbf{M}^{-1}\vec{F}_{t+\Delta t} \end{aligned} \tag{3.6}$$

où $\vec{V} = \dot{\vec{X}}$ est un vecteur 3N contenant la vitesse des N atomes dans le système, Δt est le pas de temps d'intégration et où toutes les autres variables ont la même définition qu'auparavant. Cet algorithme est le plus utilisé dans les simulations sous contraintes des positions, car il y a seulement une évaluation de la position par cycle d'intégration ce qui permet d'appliquer seulement une fois les contraintes.[62]

Le pas d'intégration Δt doit être choisi avec soin : il devrait être assez petit pour que la dynamique du système soit correctement intégrée tout en étant assez grand pour accéder plus rapidement aux échelles de temps caractéristique de la structure, de la fonction, du repliement et de l'oligomérisation des protéines. Ultimement, la grandeur de Δt est limitée par la période d'oscillation la plus courte – fréquence plus grande – dans le système. Dans le cas des protéines, la période de vibrations des liaisons covalentes contenant un hydrogène est la plus courte avec $\sim$ 10 fs. L'obtention d'une dynamique représentative nécessaire au moins environ 10 pas d'intégration durant la période de l'oscillation la plus rapide ce qui implique $\Delta t \sim$ 0,5–1,0 fs. Heureusement, les liaisons covalentes peuvent être contraintes à leur longueur d'équilibre, ce qui permet d'augmenter Δt à 1,5–2,0 fs, car les périodes d'oscillations suivantes sont plus grandes que 20 fs.[63] De plus, certains algorithmes d'intégration utilisent plusieurs pas de temps, un pour

ajoutées au système tel que présenté à la Section 3.3.

[62] Ceci est présenté plus en détail à la Section 3.3.

[63] Il s'agit des angles de valence contenant un hydrogène. Les stratégies pour contraindre les liaisons covalentes

chaque échelle de temps du mouvement. Habituellement, les forces associées aux liaisons covalentes et aux angles de valence sont intégrées plus souvent avec $\Delta t = 0{,}5$ fs, tandis que celles associées aux angles dièdres et aux interactions non-liées sont intégrées moins souvent avec $\Delta t = 1{,}0$ fs et $\Delta t = 2{,}0$ fs, respectivement. Pour ces valeurs, le ratio du nombre d'intégrations pour un cycle total est donc 4 : 2 : 1, respectivement. Les algorithmes MTS – *Multiple-timestep* – deviennent de plus en plus populaires, car (i) ils sont plus rapide computationnellement que l'application de contraintes sur les liaisons covalentes[64] et (ii) ils permettent de calculer les interactions non-liées longues portées beaucoup moins souvent avec un pas d'intégration Δt aussi long que 6 fs [140, 283].

L'algorithme Verlet vitesse dans sa forme actuelle produit une dynamique dans l'ensemble microcanonique (NVE). Il ne prend pas en compte de couplage pour maintenir la température constante (NVT) ou pour maintenir la température et la pression constantes (NPT). Pour passer de l'ensemble NVE aux ensembles NVT et NPT, l'équation du mouvement doit être modifiée telle que discutée à la Section 3.3.

La **dynamique de Langevin** est une représentation phénoménologique d'un système de particules qui subissent un amortissement et des collisions aléatoires provenant de leur environnement immédiat. C'est une formulation intéressante pour les simulations de protéines en présence implicite du solvant, car la dynamique de Langevin imite la viscosité du solvant (amortissement) et les collisions aléatoires solvant–protéine. L'équation du mouvement est

$$\begin{aligned} \dot{\vec{X}}(t) &= \vec{V}(t) \\ \mathbf{M}\dot{\vec{V}}(t) &= -\nabla E(\vec{X}(t)) - \gamma \mathbf{M}\vec{V}(t) + R(t) \end{aligned} \tag{3.7}$$

où le second terme de droite représente l'amortissement contrôlé par γ et où le troisième terme de droite est un bruit blanc caractérisé par une moyenne nulle, $\langle R(t)\rangle = 0$, et une variance donnée par $\langle R(t)R(t')^{\mathsf{T}}\rangle = 2\gamma k_B T \mathbf{M}\delta(t-t')$. La température est assurée par le dernier terme, donc il s'agit d'une simulation satisfaisant l'ensemble canonique (NVT). La valeur du paramètre de collision (γ) de chaque atome d'une protéine peut être estimée à partir de la loi de Stokes

sont abordées à la Section 3.3.

[64] Ces dernières introduisent, en effet, des corrélations entre les atomes à travers une séquence assez longue de liaisons à chaque pas de temps ce qui augmente le coût computationnel en brisant la localité du calcul.

– $\gamma = 6\pi\eta r/m$ où η est la viscosité dynamique du fluide, r est le rayon hydrodynamique de l'atome et m est la masse de l'atome – ce qui correspond à environ $\gamma \sim 50$ ps^{-1}. Habituellement, en pratique, une plus petite valeur est prise $\gamma \sim$1–5 ps^{-1} afin de ne pas trop amortir la dynamique du système et une seule valeur est utilisée pour tous les atomes.

3.2.3 La dynamique moléculaire d'échanges de répliques

La dynamique moléculaire (DM) en elle-même n'est pas efficace pour échantillonner globalement l'espace énergétique accessible à la protéine dans un temps de simulation raisonnable, car la dynamique du système peut rester prise dans des minimums locaux pendant longtemps. Plusieurs méthodes, comme la dynamique moléculaire d'échanges de répliques (REMD),[65] se basent directement sur la dynamique moléculaire, mais reformulent la situation afin d'échantillonner convenablement l'espace énergétique tout en utilisant judicieusement les ressources informatiques.

Le principe de base de REMD repose sur l'exécution de dynamiques moléculaires en parallèle à différentes conditions et l'introduction d'échanges ponctuels des coordonnées atomiques entre ces simulations [144, 322, 449].[66] La version originelle repose sur l'utilisation de simulations à différentes températures dont la température d'intérêt. Les simulations à plus hautes températures sont associées à une probabilité plus grande d'avoir des changements structuraux plus significatifs, car les hautes barrières enthapilques sont plus facilement franchissables. À l'opposé, les simulations à plus basses températures sont associées, la plupart du temps, à des changements structuraux qui sont plus locaux. Chaque simulation est appelée une réplique, laquelle diffuse d'une température et l'autre. Lors de l'étude du repliement et de l'agrégation des protéines, les températures couvrent un spectre assez large afin d'en avoir d'assez élevées pour dénaturer (déplier) la protéine. Ainsi, la protéine peut potentiellement accéder à des repliements différents lorsque la réplique re-diffusera vers les températures plus physiologiques (290–310 K).

Les échanges sont contrôlés afin que l'échantillonnage à une température donnée représente la bonne thermodynamique. En d'autres mots, à une température donnée, la fréquence d'observation de chaque structure représente ce que l'on observerait à cette température en regardant

[65] *Replica exchange molecular dynamics.*

[66] À l'origine, le protocole d'échanges de répliques avait été développé pour des simulations Monte-Carlo pour les matériaux [491], puis pour les protéines [181].

le système assez longtemps. Ceci est assuré par l'utilisation du critère de Métropolis pour calculer la probabilité que deux répliques à des températures voisines échangent leur coordonnées spatiales[67] :

$$\mathrm{P}(X_i \leftrightarrow X_j) = \min\left[1, \exp\left(-\frac{E(X_j) - E(X_i)}{k_B(T_n - T_m)}\right]\right) \tag{3.8}$$

où $\mathrm{P}(X_i \leftrightarrow X_j)$ est la probabilité d'échanger les coordonnées spatiales de la réplique i à la température T_m avec celles de la réplique j à la température T_n,[68] k_B est la constante de Boltzmann et $E(X)$ est l'énergie potentielle du système. Afin d'avoir des échanges, le critère de Métropolis exige un spectre de températures qui possède un recouvrement de l'énergie potentielle entre les températures adjacentes. La probabilité moyenne d'échanges visée est de 10–30% afin d'avoir un bon taux d'échanges tout en évitant une trop grande corrélation entre les températures adjacentes. Habituellement, une distribution exponentielle des températures est suffisante, mais il existe aussi des méthodes plus complexes qui estiment la distribution idéale [371, 386] et qui optimise à la volée la distribution de températures [466]. Les échanges sont essayés entre les répliques adjacentes avec une période fixe d'environ 1–10 ps.[69] Une fois l'échange effectué, la vitesse des atomes est renormalisée par $\sqrt{T_n/T_m}$ afin de conserver une énergie cinétique adéquate.

L'analyse des résultats se fait à la température d'intérêt, c'est-à-dire proche des températures physiologiques (290–310K). Il est aussi possible de combiner l'échantillonnage effectué à toutes les températures en pondérant adéquatement les données obtenues à différentes températures avec la méthode PT-WHAM[70] [96].

Une autre approche populaire de REMD est de laisser les répliques diffuser dans un spectre d'énergies potentielles différentes au lieu faire varier la température. Cette approche porte le nom d'échanges de répliques Hamiltonien : HREX (*Hamiltonian replica exchange*) ou H-REMD (*Hamiltonian replica exchange molecular dynamics*) [166, 448]. Il existe plusieurs façons de perturber l'énergie potentielle en échelonnant, par exemple, le rayon de van der Waals des

[67] Coordonnées des atomes du système.

[68] T_m et T_n sont des températures adjacentes.

[69] Une alternance paire-impaire/impaire-paire est introduite dans les paires de répliques utilisées afin d'éviter des conflits dans les échanges. Pour une simulation REMD avec 5 répliques, la première ronde d'échanges sera tentée entre les répliques 1 et 2 ainsi qu'entre les répliques 3 et 4, tandis que la deuxième ronde d'échanges sera tentée entre les répliques 2 et 3 ainsi qu'entre les répliques 4 et 5. Ceci est répété par alternance.

[70] *Weighted histogram analysis method for parallel tempering.*

atomes [200], l'énergie des angles dièdres ϕ et ψ [524] ou la formation des ponts hydrogènes [486]. Quant à lui, le protocole REST2 échelonne seulement les interactions non-liées électrostatiques et Lennard-Jones ainsi que les interactions liées des angles dièdres [492]. Dans ce cas, la probabilité d'un échange est contrôlée par

$$\mathrm{P}(X_i \leftrightarrow X_j) = \min\left[1, \exp\left(-\frac{E_m(X_j) - E_m(X_i) + E_n(X_i) - E_n(X_j)}{k_B T}\right)\right] \tag{3.9}$$

où $\mathrm{P}(X_i \leftrightarrow X_j)$ est la probabilité d'échanger les coordonnées spatiales[71] de la réplique i initialement à l'énergie potentielle E_m avec celles de la réplique j à l'énergie potentielle E_n,[72] k_B est la constante de Boltzmann, et T est la température. L'énergie potentielle $E(X)$ est donnée par $E_m(X) = \lambda_m E_{pp}(X) + \sqrt{\lambda_m} E_{ps}(X) + E_{ss}(X)$ où E_m est l'énergie potentielle à l'échelon m, E_{pp} est l'énergie des interactions protéine–protéine, E_{ps} est l'énergie des interactions protéine–solvant, E_{ss} est l'énergie des interactions solvant–solvant et λ_m est l'échelonnage appliqué à l'échelon m. Dans le protocole REST2, les interactions solvant–solvant ne sont donc pas échelonnées, ce qui en fait une méthode ayant un coût computationnel significativement plus faible que la méthode REMD classique. En effet, le nombre de répliques augmente très rapidement avec le nombre de degrés de liberté du système dans la méthode REMD afin de conserver un recouvrement suffisant de l'énergie potentielle entre les températures adjacentes.

Il est aussi possible de jumeler REMD et H-REMD dans un protocole hybride. Un exemple que nous avons utilisé est la méthode HT-REMD où le protocole REMD est utilisé normalement et où le protocole H-REMD est utilisé seulement à la plus haute température [256, 257]. Lors de l'étude du processus d'oligomérisation, ceci permet de séparer plus complètement les chaînes polypeptidiques afin de favoriser l'échantillonnage de nouvelles régions de l'espace structurel. Une amélioration envisageable de HT-REMD serait d'effectuer un protocole H-REMD pour lequel seules les interactions non-liées sont échelonnées. Ceci permettrait de défaire les oligomères tout en conservant une partie de la structure des monomères, ce qui respecterait plus le processus physiologique de l'agrégation : repliement de monomères suivit de leur assemblage en oligomères. Par ailleurs, ce protocole pourrait être appliqué à chaque température et non seulement à la température la plus élevée.

Finalement, il existe désormais une panoplie de protocoles REMD dont certains permettent

[71] Coordonnées des atomes du système.

[72] E_m et E_n sont des échelons adjacents.

de changer la pression, de passer d'une représentation tout-atome à une représentation gros-grain, de contrôler l'état de protonation des acides aminés, etc. Pour tous ces protocoles, le principe de base des échanges reste le même. Il est aussi possible de combiner REMD avec d'autres méthodes d'échantillonnage avancé telles que l'échantillonnage parapluie [448] et la métadynamique [19, 74, 76].

3.2.4 La métadynamique

Une autre méthode d'échantillonnage avancée est la métadynamique (MétaD) qui est disponible dans le module d'extension PLUMED pour GROMACS et NAMD ce qui en fait une méthode de plus en plus populaire [62, 467]. La MétaD repose, comme la méthode REMD, directement sur la dynamique moléculaire [19, 258, 259, 450]. Elle introduit un biais dépendant du temps qui est construit en ajoutant des distributions normales dans l'espace d'énergie aux endroits visités selon un ensemble de variables collectives (CVs) prédéterminées. Ceci augmente l'échantillonnage global tout en permettant de reconstruire le paysage d'énergie libre selon ces CVs ($\vec{S}$) puisque le potentiel de biais converge vers

$$V(\vec{S},t\to\infty) = -\frac{\Delta T}{T+\Delta T}F(\vec{S}) + C \tag{3.10}$$

où $V(\vec{S},t)$ est le potentiel de biais, $F(\vec{S})$ est l'énergie libre, T est la température et ΔT est un paramètre qui contrôle la hauteur des barrières enthalpiques échantillonnées par la méthode *well-tempered* MétaD qui assure une convergence stable de la MétaD [20, 112]. La déviation standard et la hauteur initiale des biais sont choisies selon la résolution désirée du paysage énergétique des CVs. Par ailleurs, si l'échantillonnage est suffisamment étendu, il est possible de reconstruire le paysage d'énergie libre selon n'importe quelles CVs en pondérant de façon appropriée les données [61, 465].

Pour une reconstruction adéquate du paysage d'énergie libre selon les CVs choisies, il est nécessaire de considérer toutes les CVs lentes. Or, la MétaD devient excessivement demandante computationnellement lorsqu'il y a plus de 2–3 CVs. Deux solutions existent pour résoudre cette limitation. La première consiste à effectuer des simulations MétaD en parallèle à la manière REMD où les répliques diffusent d'une simulation à l'autre et où chaque simulation est associée à une CV en particulier [381]. Cette méthode s'appelle le *Bias-exchange metadynamics* (BEMetaD). La deuxième consiste à jumeler REMD ou H-REMD directement avec la MétaD ce

qui permet d'échantillonner convenablement toutes les CVs lentes qui ne sont pas considérées par la MétaD [74, 76, 77]. Cette méthode s'appelle *parallel tempering metadynamics* PTMetaD (MétaD avec REMD) ou HREXMetaD (MétaD avec HREX) et elle sera utilisée aux Chapitres 7 et 10.

3.3 Les algorithmes auxiliaires

Plusieurs algorithmes sont associés avec les algorithmes d'échantillonnage pour, entre autres, contrôler la température et la pression, augmenter la rapidité des calculs, traiter correctement les interactions électrostatiques longues portées, et appliquer des contraintes sur la protéine. Seuls les algorithmes spécifiquement utilisés dans les prochains chapitres seront présentés.

3.3.1 Le contrôle de la température

Les thermostats Berendsen, Bussi–Donadio–Parrinello et Nosé-Hoover pour le contrôle de la température lors d'une dynamique moléculaire sont présentés dans cette section. Ils permettent de faire des simulations dans l'ensemble NVT (canonique) – nombre d'atomes, volume et température constants – en modifiant l'équation du mouvement présentée à l'équation 3.5.

Le thermostat **Berendsen** est utilisé depuis longtemps grâce à sa simplicité [40]. Les fluctuations de la température sont amorties exponentiellement selon $\frac{dT}{dt} = \frac{T_0 - T}{\tau}$ où τ contrôle l'échelle de temps du retour de T (la température actuelle) à T_0 (la température de référence). Ceci équivaut à l'ajout d'un couplage faible à la température de référence à travers les équations du mouvement selon

$$\begin{aligned} \dot{\vec{X}}(t) &= \vec{V}(t) \\ \mathbf{M}\dot{\vec{V}}(t) &= -\nabla_{\vec{X}} E(\vec{X}(t)) - \lambda_t \mathbf{M}\vec{V}(t) \quad \text{où} \quad \lambda_t = \frac{1}{2\tau_{\mathrm{T}}}\left(1 - \frac{T_0}{T(t)}\right) \end{aligned} \tag{3.11}$$

où $\mathbf{M}$ est une matrice 3N×3N contenant sur sa diagonale les masses des N atomes, $\dot{\vec{V}}$ est un vecteur 3N contenant le vecteur accélération de chaque atome, $E(\vec{X})$ est l'énergie potentielle du système, λ_t est la constante de couplage au bain thermique, et $\dot{\vec{X}}$ est un vecteur 3N contenant le vecteur vitesse de chaque atome. La simplicité de ce thermostat provient de l'absence de nouvelles variables dynamiques à l'équation du mouvement. Ceci permet d'utiliser l'intégrateur

de Verlet vitesse sans modification et de seulement échelonner la vitesse de chaque atome par $c_\mathrm{T} = \sqrt{1 - \frac{\Delta t}{\tau_\mathrm{T}}\left(1 - \frac{T_0}{T}\right)}$ où Δt est le pas de temps d'intégration, τ_T est la constante temporelle de couplage au bain thermique,[73] T_0 est la température de référence et T est la température actuelle. Pour calculer c_T, la température T doit être calculée à partir du théorème d'équipartition : $K = \sum_{i=0}^{N} \frac{1}{2} m_i v_i^2 = \frac{1}{2} N_\mathrm{dl} k_B T$ où K est l'énergie cinétique totale du système, N est le nombre d'atomes, m_i est la masse de chaque atome, v_i est le module de la vitesse de chaque atome, N_dl est le nombre de degrés de liberté du système, et k_B est la constante de Boltzmann. Le nombre de degrés de liberté lorsque le système ne subit aucune contrainte est $N_{dl} = 3N$. Or, comme dans la plupart des simulations sur les protéines, ceci n'est plus le cas lorsque la quantité de mouvement totale (3 degrés de liberté) est maintenue nulle et lorsque la longueur des liaisons covalentes est fixée à leur valeur d'équilibre. Dans ce cas, la nombre de degrés de liberté est donné par $N_{dl} = 3N - N_\mathrm{liens} - 3$ où N_liens est le nombre de liaisons covalentes contraintes. Le désavantage majeur du thermostat Berendsen est qu'il ne génère pas un ensemble canonique (NVT) correct. C'est ainsi qu'il est principalement utilisé durant la phase d'équilibration. D'autres algorithmes, décrits ci-dessous, sont utilisés pour obtenir un ensemble canonique.

Le thermostat **Bussi–Donadio–Parrinello** conserve la simplicité du thermostat de Berendsen – absence de nouvelles variables dynamiques à l'équation du mouvement – tout en générant un ensemble canonique [75]. Tout comme le thermostat de Berendsen, ce thermostat nécessite seulement de échelonner la vitesse de chaque atome par un nouveau facteur c_T ce qui requiert aucune modification à l'intégrateur de Verlet. Contrairement au thermostat de Berendsen, le thermostat Bussi–Donadio–Parrinello vise une température de référence qui est déterminée de façon stochastique à partir de la distribution canonique de l'énergie cinétique à l'équilibre. La dynamique suivie par l'énergie cinétique est ainsi décrite par

$$dK = (\bar{K} - K)\frac{dt}{\tau} + 2\sqrt{\frac{K\bar{K}}{N_{dl}}}\frac{dW}{\sqrt{\tau}} \tag{3.12}$$

où K est l'énergie cinétique actuelle, $\bar{K}$ est l'énergie cinétique de référence, τ contrôle l'échelle de temps du retour de K à $\bar{K}$, N_{dl} est le nombre de degrés de liberté, et dW est un bruit de Wiener.[74] Cette équation peut être résolue exactement afin de trouver le nouveau facteur

[73] La valeur de τ_T est choisie entre 10–100 fs. La constante de couplage au bain thermique (τ_T) est reliée à l'échelle de temps de l'amortissement de la température (τ) par $\tau = 2C_V \tau_\mathrm{T} / N_{dl} k_B$ où C_V est la chaleur spécifique à volume fixe.

[74] Cette équation vient d'une description stochastique de l'évolution de K en fonction du temps selon l'équation

(c_T) pour échelonner la vitesse de chaque atome tout en restant dans l'ensemble canonique.[75] Lorsque le terme stochastique est nul ($dW = 0$), l'équation de Berendsen pour le contrôle de la température est retrouvée : $\frac{dT}{dt} = \frac{T_0 - T}{\tau}$.

Le thermostat **Nosé-Hoover** procède par une approche complètement différente en introduisant une nouvelle variable dynamique directement dans l'équation du mouvement [188, 347]. Cette nouvelle variable dynamique agit comme un thermostat générant des fluctuations dans la quantité de mouvement des atomes pour que la dynamique du système représente l'ensemble canonique (NVT). Cette méthode, par contre, ne génère pas une dynamique ergodique, ce qui peut être partiellement corrigé en ajoutant une chaîne de variables dynamiques [308]. Ces dernières agissent séquentiellement comme thermostat du thermostat précédent dans la chaîne. Les équations du mouvement de l'ensemble de ces variables dynamiques sont pour chaque atome

$$
\begin{aligned}
\dot{\vec{X}}(t) &= \vec{V}(t) \\
\dot{\vec{V}}(t) &= -\mathbf{M}^{-1}\nabla_{\vec{X}} E(\vec{X}(t)) - \frac{p_{\xi_1}(t)}{Q_1}\vec{V}(t) \\
\dot{p}_{\xi_1}(t) &= (T(t) - T_0) - p_{\xi_1}(t)\frac{p_{\xi_2}(t)}{Q_2} \\
\dot{p}_{\xi_2}(t) &= \left[\frac{p^2_{\xi_1}(t)}{Q_1} - k_B T\right] - p_{\xi_2}(t)\frac{p_{\xi_3}(t)}{Q_3} \\
\cdots &\quad \cdots \\
\cdots &\quad \cdots \\
\dot{p}_{\xi_{N_T-1}}(t) &= \left[\frac{p^2_{\xi_{N_T-2}}(t)}{Q_{N_T-2}} - k_B T\right] - p_{\xi_{N_T-1}}(t)\frac{p_{\xi_{N_T}}(t)}{Q_{N_T}} \\
\dot{p}_{\xi_{N_T}}(t) &= \frac{p^2_{\xi_{N_T-1}}(t)}{Q_{N_T-1}} - k_B T
\end{aligned}
\tag{3.13}
$$

où p_{ξ_i} sont les quantités de mouvement associées à la coordonnée généralisée ξ_i, $d\xi_i/dt = p_{\xi_i}/Q_i$, ainsi qu'à la masse Q_i du $i^{\text{ième}}$ thermostat,[76] et où N_T est le nombre de thermostats dans

---

de Fokker-Planck : $dK = \left(D(K)\frac{\partial \log \bar{P}}{\partial K} + \frac{\partial D(K)}{\partial K}\right)dt + \sqrt{2D(K)}dW$ où la probabilité $\bar{P}(K) \propto K^{(N_{dl}/2-1)}\mathrm{e}^{-\beta K}$ et où $D(K) = 2K\bar{K}/N_{dl}\tau$ par choix.

[75] Pour le thermostat de Bussi–Donadio–Parrinello, $c_\mathrm{T}^2 = \mathrm{e}^{-\Delta t/\tau} + \frac{\bar{K}}{N_{dl}}\left(1 - \mathrm{e}^{\Delta t/\tau}\right)\left(R_1^2 + \sum_{i=2}^{N_{dl}} R_i^2\right) + 2\mathrm{e}^{-\Delta t/2\tau}\sqrt{\frac{\bar{K}}{N_{dl}K}\left(1 - \mathrm{e}^{-\Delta t/\tau}\right)}R_1$ où les R_i sont des nombres aléatoires Gaussiens.

[76] Ici, $Q = Q'/k_B N_{dl}$ où Q' est la variable utilisée dans l'article d'origine [308].

la chaîne. La température actuelle (T) est calculée à partir de l'énergie cinétique et du théorème d'équipartition : $K = \sum_{i=0}^{N} \frac{1}{2} m_i v_i^2 = \frac{1}{2} N_{\mathrm{dl}} k_B T$ où K est l'énergie cinétique totale du système, N est le nombre d'atomes, m_i est la masse de chaque atome, v_i est le module de la vitesse de chaque atome, N_{dl} est le nombre de degrés de liberté, et k_B est la constante de Boltzmann. La relaxation vers la température de référence (T_0) ainsi produite par le thermostat Nosé-Hoover est oscillatoire.[77] Plus intuitivement, la valeur de $Q = \tau_{\mathrm{T}}^2 T_0 / 4\pi^2$ est déterminée en spécifiant la fréquence moyenne de relaxation (τ_{T}) des thermostats.

Le désavantage principal des thermostats au niveau computationnel est qu'ils requièrent une communication globale de la vitesse de tous les atomes afin de calculer la température à partir du théorème d'équipartition. Ceci devient problématique dans le cas des logiciels de calculs hautes performances finement parallélisés pour la dynamique moléculaire. Une nouvelle méthode résout en bonne partie cette limitation en appliquant le formalisme de décomposition de l'intégration des équations du mouvement au thermostat de Nosé-Hoover [283]. Les variables associées au thermostat sont intégrées moins fréquemment (pas d'Intégration, $\Delta t = 30$ fs) que les autres puisque leur échelle de temps est beaucoup plus lente que celles des atomes associées aux vibrations des liaisons covalentes et des angles de valence, aux rotations des angles dièdres ainsi qu'aux interactions non-liées ($\Delta t \sim 2$ fs). La même procédure peut être appliquée pour un barostat – présenté à la prochaine section – avec $\Delta t \sim 1{,}2$ ps.

3.3.2 Le contrôle de la pression

Les barostats Berendsen et Parrinello–Rahman pour le contrôle de la pression lors d'une dynamique moléculaire sont présentés dans cette section. Couplés à un thermostat, ils permettent de faire des simulations dans l'ensemble NPT – nombre d'atomes, pression et température constants – en modifiant l'équation du mouvement présentée à l'équation 3.5.

Le barostat **Berendsen** est utilisé depuis longtemps grâce à sa simplicité [40]. Il permet d'équilibrer rapidement la pression d'un système moléculaire en amortissant exponentiellement les fluctuations de la pression selon $\frac{d\mathbf{P}}{dt} = \frac{\mathbf{P_0} - \mathbf{P}}{\tau_{\mathrm{P}}}$ où τ_{P} contrôle l'échelle de temps du retour de $\mathbf{P}$ (la pression tensorielle actuelle) à $\mathbf{P_0}$ (la pression tensorielle de référence). Le contrôle de la pression équivaut à l'ajout d'un couplage faible à la pression de référence à travers les équations du mouvement selon

[77] La relaxation est exponentielle pour les thermostats Berendsen et Bussi–Donadio–Parrinello.

$$\begin{aligned} \dot{\vec{X}}(t) &= \vec{V}(t) - \frac{\beta\left([\mathbf{P}]_{\mathbf{0}} - [\mathbf{P}]\right)}{3\tau_{\mathrm{P}}}\vec{X}(t) \\ \mathbf{M}\dot{\vec{V}}(t) &= -\nabla_{\vec{X}} E(\vec{X}(t)) \end{aligned} \tag{3.14}$$

où $\mathbf{M}$ est une matrice 3N×3N contenant sur sa diagonale les masses des N atomes, $\dot{\vec{V}}$ est un vecteur 3N contenant le vecteur accélération de chaque atome, $E(\vec{X})$ est l'énergie potentielle du système, $\dot{\vec{X}}$ est un vecteur 3N contenant le vecteur vitesse de chaque atome, τ_{P} est la constante de couplage au bain de pression et $[\mathbf{P}]$ est une matrice 3N×3N par blocs répétant N fois sur sa diagonale la matrice 3×3 $\mathbf{P}$. La simplicité de ce thermostat provient de l'absence de nouvelles variables dynamiques à l'équation du mouvement. Ceci permet d'utiliser l'intégrateur de Verlet vitesse sans modification et de seulement échelonner les vecteurs de la cellule périodique ainsi que les positions des atomes par $\mu = \mathbb{1} + \frac{\beta\Delta t}{3\tau_P}(\mathbf{P_0} - \mathbf{P})$ où $\mathbb{1}$ est une matrice identité et β est la compressibilité isotherme définie par $\beta = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)$ (ici, V est le volume).[78] Afin de calculer μ, la matrice de pression doit être déterminée par $\mathbf{P} = \frac{1}{Vol}\left[\mathbf{M}\vec{V}\vec{V}^{\mathrm{T}} + \sum_{i<j}(\vec{r}_i - \vec{r}_j)\vec{F}_{ij}^{\mathrm{T}}\right]$ où $\vec{F}_{ij}$ est le vecteur force entre les atomes i et j. Le désavantage majeur du barostat Berendsen, tout comme le thermostat Berendsen, est qu'il ne génère pas l'ensemble approprié. C'est ainsi qu'il est principalement utilisé durant la phase d'équilibration. D'autres algorithmes, décrits ci-dessous, sont utilisés pour obtenir l'ensemble approprié.

Le barostat **Parrinello-Rahman** procède en introduisant une nouvelle variable dynamique directement dans l'équation du mouvement [368]. Celle-ci agit comme un barostat en changeant les dimensions du système pour que la pression reste constante. Dans les prochains chapitres, le barostat Parrinello-Rahman est utilisé de concert avec le thermostat Nosé-Hoover afin de générer une dynamique dans l'ensemble NPT [188, 347, 348]. Les équations du mouvement pour une chaîne d'un thermostat ($N_T = 1$ dans l'équation 3.13) et pour une pression isotrope [201] sont

[78] La compressibilité isotherme est approximée par celle de l'eau à 300 K et 1 atm : 4,6·10^{-10} Pa^{-1} (1 Pa = 9,87·10^{-6} atm). La valeur précise importe peu puisque seulement le ratio β/τ_P est présent dans le calcul de μ.

$$
\begin{aligned}
\dot{\vec{X}}(t) &= \vec{V}(t) + \frac{p_\eta(t)}{m_\eta}\vec{X}(t) \\
\mathbf{M}\dot{\vec{V}}(t) &= -\nabla_{\vec{X}} E(\vec{X}(t)) - \mathbf{M}\dot{\vec{V}}(t)\left[\left(1+\frac{3}{N_{dl}}\right)\frac{p_\eta(t)}{m_\eta} + \frac{p_\xi(t)}{m_\xi}\right] \\
\dot{v}_l(t) &= 3v_l(t)\frac{p_\eta(t)}{m_\eta} \\
\dot{p}_\eta(t) &= 3v_l(t)\,(P(t)-P_0) + \frac{3}{N_{dl}}\left(2\vec{V}(t)^{\mathrm{T}}\mathbf{M}^{-1}\vec{V}(t)\right) - m_\eta p_\eta(t) p_\xi(t) \\
\dot{p}_\xi(t) &= 2\vec{V}(t)^{\mathrm{T}}\mathbf{M}^{-1}\vec{V}(t) + \frac{p_\eta(t)}{m_\eta} - (N_{dl}+1)k_B T_0
\end{aligned}
\tag{3.15}
$$

où (η, m_η, p_η) et (ξ, m_ξ, p_ξ) sont respectivement la coordonnée généralisée (η,ξ), la masse (m_η et m_ξ) et la vélocité généralisée (p_η,p_ξ) associées à la dynamique du barostat et du thermostat, N_{dl} est le nombre de degrés de liberté du système,[79] v_l est le volume de la cellule périodique, P est la pression isotrope actuelle, P_0 est la pression isotrope de référence et T_0 est la température de référence. Toutes les autres variables ont la même signification qu'auparavant. Le couplage au barostat est contrôlé par $m_\eta = \tau_{\mathrm{P}}^2(N_{dl}+3)k_B T/4\pi^2$ en spécifiant la période τ_{P}. Le couplage au thermostat est contrôlé par $m_\xi = \tau_{\mathrm{T}}^2 N_{dl} k_B T/4\pi^2$ en spécifiant la période τ_{T}. Un intégrateur populaire qui combine le thermostat Nosé-Hoover avec le barostat Parrinello-Rahman est l'algorithme Martyna–Tuckerman–Tobias–Klein (MTTK) [202].

3.3.3 Les contraintes appliquées

Dans la plupart des simulations, des contraintes sont appliquées sur la longueur des liaisons covalentes afin d'éliminer les oscillations les plus rapides dans le système. Le gain sur la rapidité des calculs est significatif : ces contraintes permettent d'utiliser un pas d'intégration d'environ 2,0 fs au lieu de seulement 0,5 fs, ce qui correspond à une multiplication de la rapidité des calculs par 4.[80] Les algorithmes qui sont utilisés dans les prochains chapitres se distinguent en trois catégories – itératif (RATTLE et SHAKE), algébrique (M-SHAKE) et analytique (LINCS et SETTLE) – ayant des domaines d'application différents.

[79] $N_{dl} = 3N - N_c$ où N est le nombre d'atomes, N_c est le nombre de contraintes appliquées sur le centre de masse et sur les molécules.

[80] Le concept de pas d'intégration a été introduit à la Section 3.2.2 conjointement avec les algorithmes d'intégration des équations du mouvement.

L'algorithme **RATTLE** est une extension directe de l'algorithme SHAKE compatible avec l'intégrateur Verlet vitesse [9, 404]. Il s'agit d'un algorithme $\mathscr{O}(N_c)$ où N_c est le nombre de contraintes et où le préfacteur est significatif. Les liaisons covalentes sont maintenues à leur longueur d'équilibre par l'ajout de contraintes appropriées directement dans l'équation du mouvement en utilisant le formalisme des multiplicateurs de Lagrange :

$$\begin{aligned} \dot{\vec{X}}(t) &= \vec{V}(t) \\ \mathbf{M}\dot{\vec{V}}(t) &= -\nabla_{\vec{X}}\left[E(\vec{X}) + \sum_{i=1}^{N_c} \lambda_i(t)\sigma_i(\vec{X})\right] \end{aligned} \tag{3.16}$$

où $\mathbf{M}$ est une matrice 3N×3N contenant sur sa diagonale les masses des N atomes, $\dot{\vec{V}}$ est un vecteur 3N contenant le vecteur accélération de chaque atome, $E(\vec{X})$ est l'énergie potentielle du système, N_c est le nombre de contraintes, λ_i sont les multiplicateurs de Lagrange et $\sigma_i = (\vec{r}_m(t) - \vec{r}_n(t))^2 - d_{mn}^2$ sont les contraintes entre toutes les paires d'atomes $\{m,n\}$ formant une liaison covalente à contraindre. Les multiplicateurs de Lagrange sont résolus itérativement pour que les contraintes soient satisfaites à la fin du pas d'intégration. Plus précisément, les étapes principales de RATTLE avec Verlet vitesse sont : (i) exécuter la première étape de l'intégrateur de l'équation 3.6 pour obtenir $\vec{V}_{t+\Delta t/2}$ à partir de $\vec{V}_t$ et de $\vec{F}_t$, (ii) calculer le vecteur décrit par chaque liaison covalente, (iii) exécuter la deuxième étape de l'intégrateur de l'équation 3.6 pour obtenir $\vec{X}_{t+\Delta t}$ à partir de $\vec{X}_t$ et de $\vec{V}_{t+\Delta t/2}$, (iv) ajuster les postions de la paire d'atomes dans chaque lien selon sa direction initiale et répéter jusqu'à l'obtention de la bonne longueur pour tous les liens,[81] (v) exécuter la dernière étape de l'intégrateur de l'équation 3.6 pour obtenir $\vec{V}_{t+\Delta t}$ à partir de $\vec{V}_{t+\Delta t/2}$ et de $\vec{F}_{t+\Delta t}$, (vi) recalculer les vecteurs de toutes les liaisons covalentes, (vi) ajuster la vitesse de la paire d'atomes dans chaque lien pour qu'elle soit nulle dans la direction du lien,[82] et (vii) débuter le prochain pas de temps. Le désavantage principal de RATTLE provient du fait qu'il s'agit d'un algorithme itératif qui doit être exécuté deux fois – une fois

[81] Cet ajustement consiste à trouver le meilleur multiplicateur de Lagrange pour satisfaire la contrainte. Il est donné par $\lambda_i = \frac{(\vec{r}_m(t+\Delta t)-\vec{r}_n(t+\Delta t))^2 - d_{mn}^2}{4\Delta t^2(m_m^{-1}+m_n^{-1})(\vec{r}_m(t)-\vec{r}_n(t))\cdot(\vec{r}_m(t+\Delta t)-\vec{r}_n(t+\Delta t))}$ où les termes quadratiques en λ sont négligés et où la contrainte de chaque lien est découplée des autres. L'itération vient justement du fait que le couplage entre les liens est négligé. Les positions optimisées sont ensuite données par $\vec{r}'_m(t+\Delta t) = \vec{r}_m(t+\Delta t) - 2(\Delta t)^2 m_m^{-1}\lambda_i(\vec{r}_m(t)-\vec{r}_n(t))$ avec $\lambda_i \to -\lambda_i$ pour $m \to n$.

[82] Ceci est nécessaire, car la dérivée de la contrainte σ_i par rapport au temps implique $(\vec{r}_m(t) - \vec{r}_n(t)) \cdot (\vec{v}_m(t) - \vec{v}_n(t)) = 0$.

pour les positions et une fois pour les vitesses – à chaque pas d'intégration. Dans le cas d'un logiciel de dynamique moléculaire finement parallélisé, il possède une limitation importante supplémentaire : les forces, vitesses et positions de tous les atomes doivent être amenées sur le même processeur pour que l'algorithme puisse itérer sur tous les liaisons covalentes.

L'algorithme **M-SHAKE** est une version analytique de SHAKE qui résout les équations de contraintes directement par inversion de matrice, sans négliger le couplage entre les liens [251].[83] L'inversion matricielle utilise l'algorithme de décomposition LU qui est d'ordre $\mathscr{O}(N_c^3)$, comparé à $\mathscr{O}(N_c)$ pour SHAKE standard, où N_c est le nombre de contraintes.[84] À première vu, M-SHAKE peut sembler beaucoup plus lent que ce dernier, mais il est significativement plus rapide que SHAKE pour les petits systèmes de 3–10 contraintes. Ceci est dû au préfacteur du $\mathscr{O}$ qui est significativement plus élevé pour SHAKE dû à sa nature itérative. M-SHAKE est couramment utilisé, entre autres, pour contraindre seulement les liaisons covalentes dans lesquelles participe un hydrogène. Contraindre seulement ces liaisons covalentes avec M-SHAKE semble être suffisant pour pouvoir utiliser un pas d'intégration minimal de $\Delta t = 2$ fs [283].

L'algorithme **LINCS** utilise aussi l'approche des multiplicateurs de Lagrange [184]. Ceux-ci sont résolus approximativement et directement dans les équations du mouvement en considérant seulement un nombre limité de couplages entre les liaisons covalentes. Pour ne requérir qu'une seule itération afin d'atteindre une convergence adéquate, une liaison covalente est couplée à toutes les autres liaisons à moins de quatre liaisons de distance. Dans le cas des protéines, le couplage à plus longue portée est en effet négligeable. Les étapes sont : (i) mise à jour de la position des atomes par l'intégrateur, (ii) application des contraintes dans la direction originelle de la liaison covalente et (iii) correction de l'effet de rotation dû au fait que la contrainte est appliquée dans le direction originelle de la liaison. Il s'agit d'un algorithme assez populaire avec GROMACS, car il permet de contraindre toutes les liaisons covalentes plus rapidement et localement en mémoire que RATTLE.

L'algorithme **SETTLE** contraint spécifiquement la géométrie des molécules d'eau ($N_c = 3$) selon une approche analytique [326]. En effet, les modèles des molécules d'eau ont, pour la plupart, été développés avec une géométrie rigide qui doit être conservée durant une simulation

[83] L'approximation linéaire qui consiste à ne considérer que les termes d'ordres $\mathscr{O}(\lambda)$ et plus petits est aussi appliquée dans M-SHAKE.

[84] L'équation algébrique à résoudre est $[\mathbf{A}]\vec{\lambda} = \vec{c}$ où $c_l = \frac{(\vec{r}_m(t+\Delta t)-\vec{r}_n(t+\Delta t))^2 - d_{mn}^2}{4\Delta t^2}$ et $A_{ll'} = \left[m_m^{-1}(\delta_{m,m'} - \delta_{m,n'}) + m_n^{-1}(\delta_{n,n'} - \delta_{n,m'})\right] \quad (\vec{r}_{m'}(t) - \vec{r}_{n'}(t)) \cdot (\vec{r}_m(t+\Delta t) - \vec{r}_n(t+\Delta t))$ où la même notation que SHAKE est utilisée.

de dynamique moléculaire. Le principe est d'illustrer la molécule d'eau comme un triangle dont les côtés sont formés par les deux liens H–O et un lien fictif H–H. Il existe une transformation unique entre le triangle idéal voulu et le triangle issu de l'intégration des équations du mouvement sans contrainte pour la position des trois atomes. Cette transformation peut être exprimée selon trois angles, lesquels sont uniques. SETTLE résout donc analytiquement ces trois angles au lieu d'utiliser les multiplicateurs de Lagrange. La contrainte des vitesses (voir paragraphe sur RATTLE) doit cependant être résolue par la solution algébrique de 3 équations utilisant les multiplicateurs de Lagrange.

CHAPITRE 4

PRÉAMBULE AUX ÉTUDES SUR LA PROTÉINE Aβ

Les rôles physiologiques de la protéine bêta-amyloïde (Aβ) ne sont pas très bien définis. La protéine Aβ est, au contraire, connue pour son rôle prépondérant dans le développement et l'évolution de la maladie d'Alzheimer [179, 489]. Il s'agit d'une maladie neurodégénérative qui possède des effets personnels dévastateurs et irréversibles dus à une perte progressive des neurones du cortex cérébral et de certaines régions subcorticales menant à l'atrophie du cerveau. Il s'en suit des pertes de mémoire de plus en plus importantes jusqu'à une perte d'autonomie complète et, ultimement, la mort. Des estimations faites par l'association *Alzheimer's disease international* montrent que la maladie d'Alzheimer, avec le vieillissement de la population, touchera mondialement plus de 100 millions de personnes et coûtera plus de mille milliards de dollars en soins prodigués annuellement à l'horizon de 2050 [506].

La caractéristique principale de cette maladie est la formation dans les tissus nerveux de plaques extracellulaires composées de protéines Aβ et de neurofibrilles intracellulaires composées de protéines Tau hyperphosphorylées [488]. Dans les plaques extracellulaires se trouvent des structures appelées fibres amyloïdes qui contiennent plusieurs milliers de protéines Aβ dans un arrangement très ordonné et riche en feuillets-β. L'origine exacte de la neurotoxicité par Aβ dans la maladie d'Alzheimer reste encore à élucider. Il est cependant reconnu que les fibres amyloïdes insolubles qui composent les plaques extracellulaires ne seraient pas directement responsables de la neurotoxicité. Ce sont les petits agrégats solubles formés de quelques protéines Aβ qui corrèlent avec la neurotoxicité en affectant la structure et la plasticité des neurones par divers mécanismes [37, 179, 246, 296, 418, 489]. Des assemblages aussi petits que deux protéines Aβ (dimères) sont neurotoxiques et sont présents dans le cortex de cerveaux atteints par la maladie d'Alzheimer [422]. Le dimère est aussi crucial dans la formation de plus gros agrégats [45].

Plusieurs approches thérapeutiques envisagent donc d'empêcher la formation de ces oligomères neurotoxiques. Or, la recherche de ces composés est complexifiée par l'existence d'un ensemble varié d'expressions physiologiques de Aβ ayant des longueurs variant entre 39 à 43 acides aminés qui possède des propriétés biochimiques distinctes. La protéine Aβ est, en effet,

produite en différentes longueurs lors de sa synthèse à partir du clivage de la protéine précurseur de l'amyloïde (APP) – une protéine membranaire d'environ 700 acides aminés – par l'enzyme β-sécrétase, puis par l'enzyme γ-sécrétase. La séquence d'acides aminés des formes physiologiques étudiées dans cette thèse est illustrée à la Figure 4.1. La forme la plus abondante, en général, est Aβ_{40}, tandis que Aβ_{42} est la forme la plus toxique, la plus encline à s'agréger et la plus abondante dans les plaques extracellulaires [179]. Certaines mutations peuvent aussi être présentes dans la séquence de Aβ comme la mutation Iowa, Aβ_{40}(D23N), qui s'agrège et forme des fibres amyloïdes différemment de Aβ_{40} et Aβ_{42} [471] et qui provoque une forme sévère de la maladie d'Alzheimer [172].

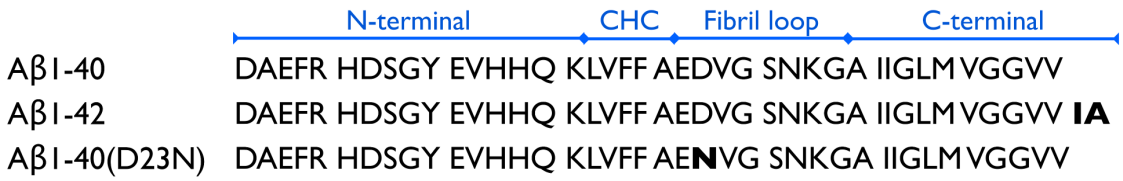


Figure 4.1 – Illustration de la séquence d'acides aminés des formes physiologiques de la protéine Aβ étudiées. La nomenclature des segments – N-terminal, CHC, *fibril loop* et C-terminal – utilisée lors des discussions sur Aβ est aussi présentée.

À ce jour, il n'existe aucun médicament qui guérit la maladie d'Alzheimer ; seuls cinq médicaments, qui ne visent pas directement Aβ, ralentissent la progression de la maladie si elle n'est pas à un stade trop avancé. Afin d'aider la découverte de composés plus efficaces qui visent directement la protéine Aβ, il est crucial de comprendre les mécanismes menant à la formation des petits agrégats (oligomères) et des fibres amyloïdes. Il devient alors nécessaire d'avoir une évaluation des interactions dominantes lors des processus d'agrégation de ses diverses formes physiologiques.

Plusieurs expériences ont dévoilé la cinétique de formation des fibres amyloïdes par microscopie à fluorescence ainsi que leurs structures finales par microscopie électronique (ME), diffraction de rayons-X (DRX) et résonance magnétique nucléaire (RMN) [469, 470]. Par ailleurs, la cinétique de formation des oligomères ainsi que leurs propriétés structurelles globales – population globale de structures secondaires et forme tridimensionnelle globale – ont été mesurées pour plusieurs formes physiologiques par diverses techniques telles que le dichroïsme circulaire (DC), le *photo-induced cross-linking of unmodified protein* (PICUP) suivi par de l'électrophorèse sur gel (SDS-PAGE) et la mobilité ionique couplée à la spectroscopie de masse

(IM-MS) [45, 53, 55, 169]. Ces études mettent en évidence les régions importantes lors de l'agrégation de Aβ en comparant ses formes physiologiques. Par exemple, la présence de deux acides aminés non-polaires au C-terminal de Aβ_{42} (Aβ_{40} vs. Aβ_{42}) change complètement la population [53, 55, 93] et la forme des oligomères [45].

Le monomère quant à lui adopte un ensemble de structures désordonnées [277, 511, 528] qui fluctuent significativement [523] et qui adoptent plusieurs coudes [265]. De plus, des résultats obtenus par PICUP et DC suggèrent que le dimère est plus structuré que le monomère pour Aβ_{40} [356]. Or, très peu de détails atomiques sont connus sur les monomère et les oligomères de Aβ. Ils sont difficilement observables en utilisant des techniques expérimentales à haute résolution comme la RMN, car ils s'agrègent rapidement et ils forment des ensembles hétérogènes de structures tridimensionnelles. Les chemins d'agrégation de Aβ sont, en effet, très complexes (Figure 4.2). Le monomère adopte un ensemble de structures variées qui s'agrègent en oligomères ayant eux aussi des structures variées qui peuvent ensuite s'assembler en fibres amyloïdes [179, 417, 489]. Toutes ces structures peuvent aussi interagir avec une membrane lipidique perturbant son intégrité et elles peuvent mener à la formation de pores transmembranaires [263, 264].

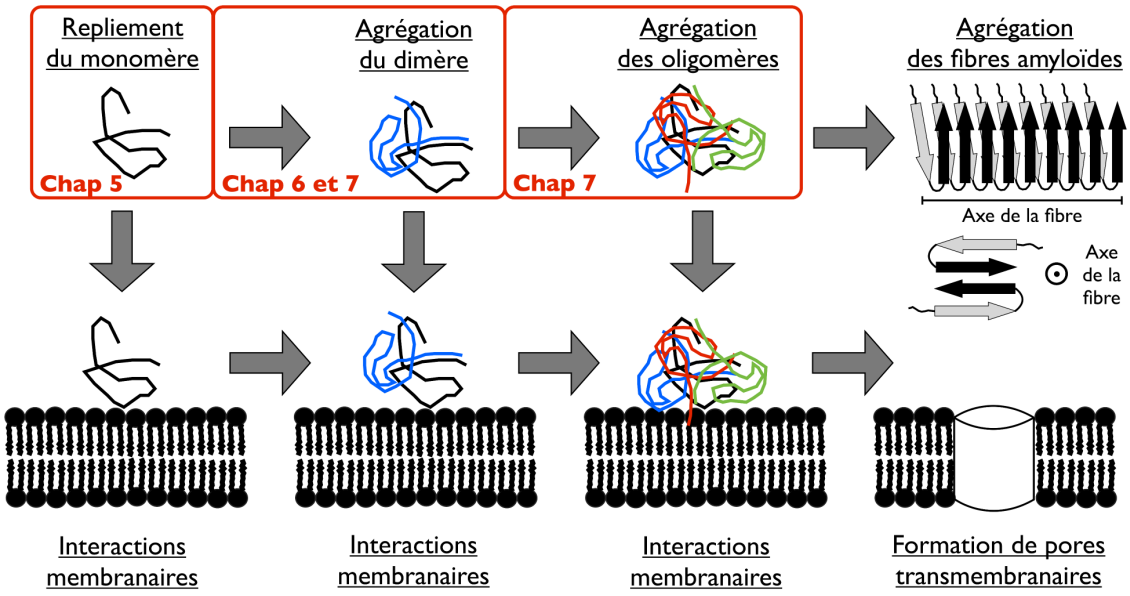


Figure 4.2 – Illustration des chemins d'agrégation de la protéine Aβ. Les systèmes étudiés aux Chapitres 5, 6 et 7 sont indiqués.

La complexité des chemins d'agrégation des petits oligomères peut être décrite au niveau

atomique par des expériences *in silico* (par ordinateur) pour compléter les données expérimentales. Dans ce contexte, nous avons décidé de simuler l'auto-assemblage du dimère de trois formes physiologiques de la protéine Aβ : Aβ_{40}, Aβ_{42} et Aβ_{40}(D23N). Les formes Aβ_{40} et Aβ_{42} ont été sélectionnées puisqu'elles sont les plus abondantes physiologiquement, et la forme Aβ_{40}(D23N) parce que ses chemins d'agrégation sont très différents des deux autres formes et parce qu'elle contient la mutation (D23N) d'un acide aminé ayant un rôle important durant l'agrégation et la formation de la fibre (Figure 4.1). Au début de ces simulations, à l'été 2009, les résultats de plusieurs simulations sur des fragments de monomère avaient déjà été publiés, mais il n'y avait aucune simulation sur le dimère. De plus, seulement deux simulations sur la séquence complète existaient pour le monomère [420, 512]. Nous avons donc choisi de caractériser aussi le repliement du monomère avec la même technique qui est utilisée pour le dimère afin de pouvoir comparer nos résultats directement avec ceux de ces simulations et ceux des expériences.

Les interactions dans nos simulations sont modélisées avec le champ de force gros-grain OPEPv3 [311]. La philosophie derrière ce champ de force, sa fonction d'énergie ainsi que ses étapes de développement sont décrites à la Section 3.1.4. Ce choix a été motivé par les raisons suivantes : (i) OPEP génère un ensemble structurel pour le fragment Aβ_{21-30} qui est en accord avec les résultats RMN [92], (ii) OPEP replie des protéines variées par plusieurs méthodes comme la technique d'activation-relaxation (ART) [332, 442, 494, 496], la REMD [86] et la reconstruction de novo [310, 311], (iii) OPEP-REMD génère l'ensemble structurel de peptides désordonnés en accord avec l'expérience à des températures physiologiques [86] et (iv) l'emploi d'un modèle gros-grain diminue le coût computationnel de la méthode REMD.

L'échantillonnage est effectué par la méthode HT-REMD, qui ajoute un protocole d'échanges de répliques Hamiltonien à la température la plus élevée d'un protocole REMD standard. Ces méthodes sont décrites à la Section 3.2.3. Nous l'avons privilégiée afin de favoriser un échantillonnage plus efficace en permettant une dissociation plus complète du dimère. La même méthode a été utilisée pour le monomère afin de permettre une comparaison directe avec les résultats du dimère, malgré le fait que la méthode REMD aurait probablement fourni un échantillonnage très suffisant en soit.

Nos résultats sur le repliement du monomère et l'auto-assemblage du dimère de Aβ_{40}, Aβ_{42} et Aβ_{40}(D23N) sont respectivement présentés aux Chapitres 5 et 6. Depuis que nous avons

entrepris et publiés nos simulations, les résultats d'autres simulations sur le monomère et le dimère de diverses formes physiologiques ont été publiés par plusieurs groupes de recherche. Nous avons donc effectué une comparaison détaillée, présentée au Chapitre 7, des résultats pour le dimère et les petits oligomères afin d'extraire des tendances relatives aux interactions cruciales et universelles lors des premières étapes d'agrégation de la protéine Aβ.

CHAPITRE 5

MONOMÈRES Aβ_{1-40}, Aβ_{1-42} ET Aβ_{1-40}(D23N)

Distinct Morphologies for Amyloid Beta Protein Monomer : Aβ_{1-40}, Aβ_{1-42} and Aβ_{1-40}(D23N)

Sébastien Côté†, Philippe Derreumaux‡, and Normand Mousseau†

† Département de Physique and Groupe d'étude des protéines membranaires (GÉPROM), Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec), Canada

‡ Laboratoire de Biochimie Théorique, UPR 9080 CNRS, Institut de Biologie Physico Chimique, Institut Universitaire de France, Université Paris Didero - Paris 7, 13 rue Pierre et Marie Curie, 75005 Paris, France

Article paru : Côté, S., P. Derreumaux, and N. Mousseau. 2011. Distinct morphologies for amyloid beta protein monomer : Aβ_{1-40}, Aβ_{1-42} and Aβ_{1-40}(D23N). *J. Chem. Phys. B.* 7 :2584-2592.

5.1 Abstract

Numerous experimental studies indicate that amyloid beta protein (Aβ) oligomers as small as dimers trigger Alzheimer's disease. Precise solution conformation of Aβ monomer is missing since it is highly dynamic and aggregation prone. Such a knowledge is however crucial to design drugs inhibiting oligomers and fibril formation. Here, we determine the equilibrium structures of the Aβ_{1-40}, Aβ_{1-42} and Aβ_{1-40}(D23N) monomers using an accurate coarse-grained force field coupled to Hamiltonian-temperature replica exchange molecular dynamics simulations. We observe that even if these three alloforms are mostly disordered at the monomeric level, in agreement with experiments and previous simulations on Aβ_{1-40} and Aβ_{1-42}, striking morphological differences exist. For instance, Aβ_{1-42} and Aβ_{1-40}(D23N) have higher β-strand propensities at the C-terminal, residues 30-42, than Aβ_{1-40}. The D23N mutation enhances the conformational freedom of the residues 22-29 and the propensity for turns and β-strands in the other regions. It also changes the network of contacts, the N-terminal (residues 1-16) becoming more independent from the rest of the protein, leading to a less compact morphology than the wild-type sequence. These structural properties could explain in part why the kinetics and the final amyloid products vary so extensively between the Aβ_{1-40} and the Aβ_{1-40}(D23N) peptides.

5.2 Introduction

Alzheimer's disease (AD) is characterized by the presence of extracellular neuritic plaques and intracellular neurofibrillary tangles in the brain [417]. Senile plaques are made of the amyloid beta (Aβ) protein. This protein is naturally produced through the cleavage of the amyloid precursor glycoprotein (APP) by the $\beta-$ and $\gamma-$secretases. Many alloforms with amino acid lengths varying between 39 and 43 are produced. Of these, Aβ_{40} is the most abundant and Aβ_{42} is the most toxic and aggregation prone [179].

Experimental studies using CD spectroscopy and electron microscopy [243] indicate that Aβ peptides exhibit a transition from random coil to β-sheet during fibrillation. Other experiments reveal that the final amyloid product, the nucleation and elongation rates [354] are distinct for different alloforms and are modulated by experimental conditions [376]. The oligomerization pathways also vary for different alloforms [53, 93] and Aβ oligomers, rather than the amyloid fibrils, are the principal neurotoxic agents [179, 293] interacting with receptors [406], metal-ions [152], cell membrane [59] and synapses [418]. Despite these significant advances in Aβ amyloid fibril assembly, precise aqueous solution conformations of Aβ monomer is missing. We know from solution NMR that it is mostly coil-turn with little β-strand content in water solution [277, 528], that Aβ_{42} is more structured at the C-terminal than Aβ_{40} [511] and oxidation of Met35 causes important changes in the monomer structure [192]. Also, hydrogen/deuterium exchange experiments showed that the monomer is completely exposed to the solvent, revealing that it is highly fluctuating [523]. Finally, limited proteolysis/mass spectrometry on Aβ_{40} monomer suggests many turns such as between Val24 and Lys28 which was observed by solution NMR on the Aβ_{21-30} fragment [265].

Yet detailed conformational knowledge of Aβ monomer is of utmost importance for two reasons. Firstly, the monomer exists in equilibrium with oligomers during polymerization [53], interacts with cell membrane [59], binds to fibril [148] and may even be neuroprotective [170]. Secondly, since Aβ oligomers as small as dimers have been implicated in AD [422], it is important to block oligomerization at the monomer level. To complement biophysical studies, computer simulations are often used. Several molecular dynamics (MD) simulations have already been reported on Aβ_{40} [510] and Aβ_{42} [157] folding in aqueous solution. Other studies examined the role of ion-binding [388], oxidation of Met35 [468] and salt-bridge Asp23-Lys28 [394] on folding of Aβ monomers. These simulations revealed many interesting features, but the gene-

rated conformational ensemble is affected by the starting structure used and the short timescale explored. Moreover, MD simulations easily get trapped in local minima biasing the results when the energy landscape is rugged as is the case with Aβ.

Replica exchange molecular dynamics (REMD) simulation, which enhances conformational sampling [449], has also been used. For the monomer of Aβ_{40} and Aβ_{42}, it was coupled to different all-atom force fields with explicit [420] and implicit solvent models [512]. Each replica was however simulated for 60-100 ns [420, 512], which may not be enough to sample the complete morphological ensemble. This convergence problem certainly holds for the simulation of Aβ_{1-39} [8]. A recent extensive all-atom simulation with implicit solvent on the truncated Aβ_{10-40} monomer reached convergence [453]. However, the predicted percentage of α-helix content ($\sim$ 38%) is significantly overestimated compared to CD analysis (between 9 and 12% [243, 355]) and the weak α-helix signal detected by NMR [528]. In addition, the N-terminal segment must not be discarded as it has been recently recognized that mutations at positions 6 and 7 alter monomer folding and oligomerization [356]. Finally, two long all-atom simulations based on different force fields and sampling approaches were recently performed on the Aβ_{42} monomer. They reported, however, very different conformational ensembles : with high β-sheet content and forming for instance a four-stranded anti-parallel β-sheet using Monte Carlo simulated annealing and an implicit solvent [324] or an ensemble of predominantly random coil conformations from a 225 ns REMD simulation using AMBER force field with explicit solvent [419].

These divergent computational conclusions on the nature of the conformations that characterize the whole ensemble of Aβ_{40} and Aβ_{42} monomer motivated us to re-explore their foldings using a different approach. Here, we use the coarse-grained protein force field (OPEP), which has been tested on widely different systems [311], and we couple it to an hybrid of hamiltonian and temperature replica exchange molecular dynamics (HT-REMD) [256], which allows a more efficient sampling of the conformational space than standard temperature REMD. Along the wild-type Aβ_{40} and Aβ_{42} peptides, we also examined the Aβ_{40}(D23N) variant implicated in early onset *AD* [172]. To our knowledge, there is no experimental and computational study on the Aβ_{40}(D23N) monomer. Yet, the mutation D23N has strong consequences on polymerization and fibril product. While the kinetics of Aβ_{40} and Aβ_{42} polymerization shows a lag phase, with Aβ_{42} peptide forming fibrils at a much faster rate, the Aβ_{40}(D23N) peptide does not display any lag

phase. [471]. In addition, Aβ_{40} and Aβ_{42} form in-register parallel β-sheet fibrils [294, 375, 469] whereas Aβ_{40}(D23N) predominant fibrils form anti-parallel β-sheets [471]. Such fibrillar morphologies are usually seen only for short sequences making therefore Aβ_{40}(D23N) a very special alloform.

5.3 Methodology

We use the implicit coarse-grained potential OPEP 3.2 parameter set [311] coupled with HT-REMD [256] to describe the morphologies visited by Aβ_{1-40}, Aβ_{1-42} and Aβ_{1-40} (D23N) alloforms using the same protocol. Aβ_{40} has the following amino acid sequence : DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV. For Aβ_{42}, two more residues, Ile41 and Ala42, are present at the C-terminal end and, for Aβ_{40}(D23N), the positively charged residue Asp23 is mutated into the neutral Asn23. We use 23 temperatures following an exponential distribution : 200.0, 229.7, 239.9, 259.4, 261.5, 273.1, 285.1, 297.7, 310.8, 324.6, 338.9, 353.9, 369.5, 385.8, 402.8, 420.6, 439.2, 458.6, 478.6, 500.0, 502.0, 504.0 and 505.0 K. At the highest temperature, we use a potential reduction scale composed of 5 steps : 0.8, 0.7, 0.6, 0.4 and 0.2 that fractionally reduces non-bonded attractive forces. Exchanges between neighboring replicas are attempted every 7.5 ps resulting in an exchange rate of $50-60\%$. This relatively high exchange rate is due to the use of a coarse-grained protein model coupled with an implicit solvent, which decreases the effective number of degrees of freedom, and to the low secondary structure probability of the peptides as discussed in Section 5.4. In the absence of clear structural differences, configurational energies between nearby temperatures tend to be closer.

Bond lengths are constrained with the RATTLE algorithm [9]. The simulations are thermalized using Berendsen's thermostat with a coupling constant of 0.1 ps [40] and have an integration time step of 1.5 fs. Simulations are started from a random extended conformation with an end-to-end distance of 36.9 Å, 37.5 Å and 36.9 Å for Aβ_{1-40}, Aβ_{1-42} and Aβ_{1-40}(D23N) respectively. Each monomer is simulated in a sphere of 60 Å-radius with reflecting boundary conditions. The size chosen for the sphere minimizes boundary conditions bias. Each replica of Aβ_{1-40} is simulated for 700 ns (giving 19.6 μs total time), each replica of Aβ_{1-42} for 900 ns (giving 25.2μs total time) and each replica of Aβ_{1-40}(D23N) for 900 ns (giving 25.2μs total time). These extensive simulations are necessary to reach convergence as discussed below.

Potential. We choose the implicit solvent coarse-grained potential OPEP 3.2, Optimized Po-

tential for Efficient structure Prediction, because it captures the main interactions during protein folding without costing much computational time as would have an all-atom explicit solvent potential [311]. This potential has been shown to recover the native structure of a variety of peptides with widely different secondary and tertiary structures as accurately as all-atom potentials using temperature REMD simulations [86]. OPEP has also been applied to short and long amyloid sequences giving results with strong similarities with experiments [88, 92, 138, 255, 256, 274, 290]. Coupled to a greedy algorithm and a structural alphabet, OPEP was able to locate, using a benchmark of 25 peptides with 9-23 amino acids, lowest energy conformations differing by 2.6 Å $C\alpha$ root mean square deviation (RMSD) from the full NMR structures [309, 310]. Briefly, this model approximates each amino acid by 6 beads : N, H_N, C_α, SC, C and O where the side chain (SC) is represented by a unique bead with glycines having an H instead of SC. The interaction parameters are finely tuned against protein structures and thermodynamics, and include bond-lengths, bond angles, improper torsions, dihedral angles, van der Waals interactions and two-body and four-body cooperative hydrogen bonds.

HT-REMD. In order to determine the conformations of Aβ, we use an hybrid of Hamiltonian and temperature replica exchange molecular dynamics (HT-REMD). T-REMD is widely used to simulate protein aggregation with a variety of implicit and explicit solvents [449]. Nonetheless, for some proteins, it is observed that T-REMD alone is not sufficient to completely unfold the protein and escape from strong local minima. HT-REMD enhances sampling by reducing non-bonded attractive forces at the highest temperature [256]. This allows Aβ to unfold completely into an extended chain.

As T-REMD [86, 449], HT-REMD has a distribution of temperatures allowing exchanges between neighboring replicas i and j according to the Metropolis criterion that preserves thermodynamic ensembles :

$$Prob(i \leftrightarrow j) = \min\left\{1, \exp\left[\frac{E_i - E_j}{k_B\left(T_i - T_j\right)}\right]\right\} \tag{5.1}$$

where $Prob(i \leftrightarrow j)$ is the probability of exchange between replicas i and j, E_i and E_j are the energies of replicas i and j respectively and T is the temperature.

As H-REMD [166], HT-REMD possesses an energy scale at the highest temperature in which non-bonded attractive forces are fractionally reduced. Exchanges between neighboring replicas i and j are governed by :

$$Prob(i \leftrightarrow j) = \min\left\{1, exp\left[-\frac{H_i(X') - H_i(X) + H_j(X) - H_j(X')}{k_B T_{max}}\right]\right\} \tag{5.2}$$

where H denotes the hamiltonian (energy) and X and X' are the configurations of replicas i and j respectively. By decreasing the non-bonded attractions between atoms at the highest temperature, HT-REMD helps the replicas escape from deep local minima increasing the conformational space sampling.

Analysis. We analyze the secondary and tertiary structures at 325 K as justified in Section 5.4.2 using the equilibrated time interval of each alloform. The secondary structure is analyzed using STRIDE [164] and the tertiary structure is analyzed using contacts between side chains and clustering. Contacts are considered when the distance between two side chain beads is smaller than the sum of their van der Waals radii plus 0.5 Å. For clustering, we first calculate the RMSD between all structures for each alloform. We then find the biggest cluster, remove from the pool of structures all those contained in this cluster and repeat iteratively until no structure is left [119]. We select a small threshold of 2 Å C_α-RMSD to increase the discrimination between clusters.

The weighted histogram analysis method [96] is used to calculate the free energy and the entropy is calculated using $S = (E - F)/T$. J-coupling constants between H_N and H_α, ${}^3J_{H_\alpha H_N}$, are calculated using the Karplus equation [225] with three different sets of coefficients [412, 419, 487]. Error bars for the J-coupling constants and the secondary structure propensities show the interval of confidence on the mean value given by the bootstrap statistical analysis method [343].

Convergence. Convergence in our simulations is assessed by three criteria. First, we check that, at equilibrium and based on the ergodic principle, the entropy as a function of temperature, $S(T)$, is time independent when averaged over sufficiently long periods as is seen in Figure 5.1.

Aβ_{40} is the most rapidly converged simulation, $S(T)$ converges within 200 ns time intervals after only 100 ns of equilibration. Aβ_{42} converges more slowly, the equilibration takes 300 ns and the entropy becomes time independent when averaged over 300 ns time windows. Aβ_{40}(D23N) requires an equilibration time of 500 ns and $S(T)$ remains constant within 200 ns time windows. Note that the differences in the time windows to reach equilibrium and obtain converged entropies starting from similar conformations for the three alloforms cannot be related to differences in aggregation properties. They only indicate that the configuration space of Aβ_{42} is more complex than those of Aβ_{40} and Aβ_{40}(D23N). We further note that the entropy of the

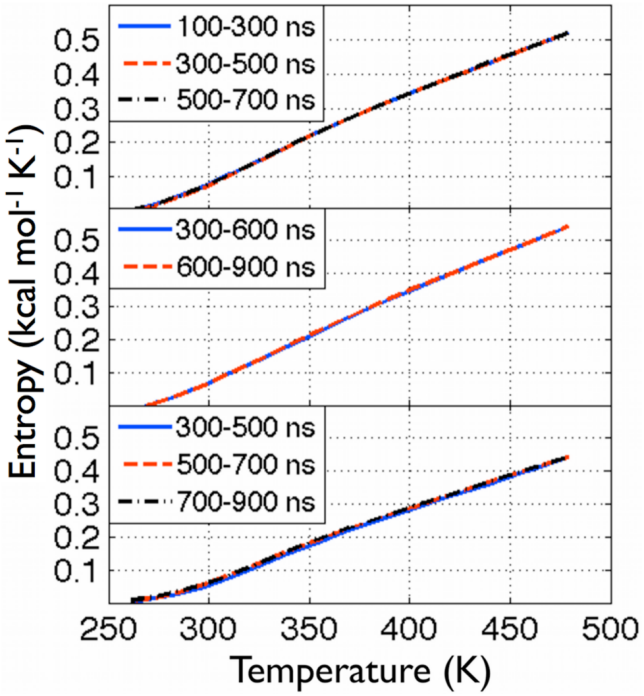


Figure 5.1 – Entropy as a function of temperature $S(T)$ averaged over different time intervals. Aβ_{40}, Aβ_{42} and Aβ_{40}(D23N) are shown in the top, middle and bottom panels respectively.

three alloforms have a similar slope, which is related to their weak secondary structure signals and their overall behavior as random coil polymers. Second, to further confirm convergence, we look at the variation of secondary structure using the time windows mentioned above for each alloform. The same trend is observed in each time window and the probabilities of secondary structures per residue vary on average by only 4-5% implying convergence (data not shown). Third, as a final check, we cluster all structures in each time window for each alloform and find that the resulting distribution of clusters are very similar between different time windows. These results confirm that, with our simulation protocol, each alloform has fully converged. As a result, for analysis, we use the equilibrated time interval of each alloform : from 100 to 700 ns for Aβ_{40}, from 300 to 900 ns for Aβ_{42} and from 500 to 900 ns for Aβ_{40}(D23N) totalizing 80 000, 80 000 and 53 333 structures respectively.

Naming convention. To facilitate analysis, the sequence is often split into four regions : the N-terminal (residues 1-16), the central hydrophobic core or CHC (residues 17-21), the loop region (residues 22-29) and the C-terminal (residues 30-42). The N-terminal is mostly hydrophilic, the CHC corresponds to ^{17}LVFFA^{21}, the loop region refers to the residues forming a loop

in the fibrillar morphologies and the C-terminus is mostly hydrophobic.

5.4 Results

5.4.1 J-couplings

Before comparing the properties of the three alloforms, it is important to select the appropriate simulation temperature for comparison with experiments. It is known that even all-atom simulations with explicit solvent generally show a shift between the predicted and experimentally observed melting temperatures, indicating that the comparison between computation and experiments cannot be made directly with the temperature used in the experiments[427]. This feature has also been observed with the OPEP force field [88]. To find the simulation temperature corresponding best to experiments, we calculate the J-coupling constants, $^{3}J_{H_N H_\alpha}$, of Aβ_{40} and Aβ_{42} at 300 K and 325 K (the highest temperature before transition to random coil), and compare those with experimental J-coupling constants measured at 300 K [420]. We also examine the use of three sets of parameters for the Karplus equation (namely the Vuister's [487], Sgourakis' [419] and Schmidt's [412] parameter sets) in calculating the J-coupling constants from the simulations.

Figure 5.2 superposes the J-coupling constants calculated for both alloforms at 300 K and 325 K using the Vuister's parameter set [487] on the experimental values [420]. In inset, the Vuister's parameter set is compared to that proposed recently by Sgourakis *et al.* [419] for the simulations at 325 K. Qualitatively, both temperature data sets follow the same trend for Aβ_{40} and Aβ_{42}. We note, however, for the Vuister's set, an overall J-coupling shift in the N-terminal region for Aβ_{40} compared to experiment, a shift that is considerably reduced with Sgourakis' set. Quantitatively (Table 5.I), the Pearson correlation coefficient (PCC) remains nearly the same for Aβ_{40} independently of the simulation temperature and the parameter set used, but is significantly improved for Aβ_{42} at 325 K, indicating that this latter temperature is most relevant for comparison with experiment.

Using the same residues and experimental J-coupling constant values [420] as benchmark, our results can be compared to the PCC obtained by other computational studies. For Aβ_{42}, our PCC value of 0.43 agrees with the values derived from all-atom REMD simulations in explicit solvent using OPLS (PCC of 0.43 using 60 ns per replica) and AMBER99SB (PCC of 0.4-0.5 using 225 ns per replica) force fields [419, 420]. For Aβ_{40}, our PCC value of 0.29 is lower than

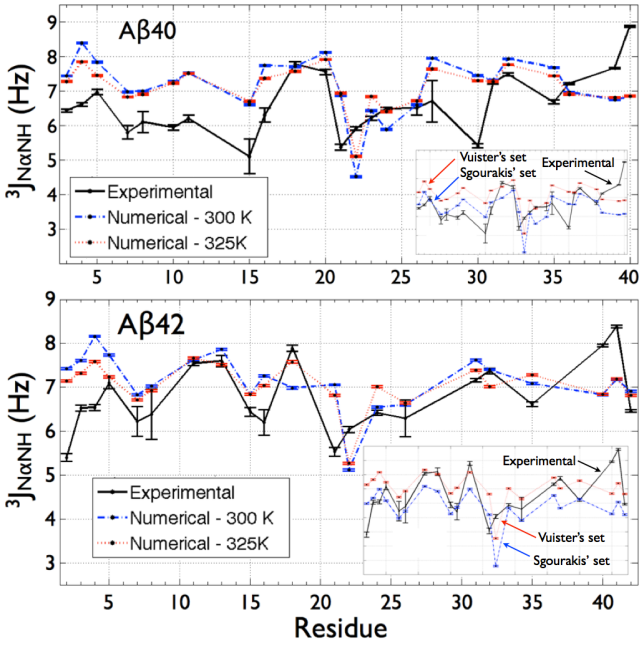


Figure 5.2 – J-coupling constants $^3J_{H_\alpha H_N}$ for Aβ_{40} and Aβ_{42} at 300K and 325K. The J-coupling constants determined experimentally [420] for Aβ_{40} and Aβ_{42} are compared to the predicted J-coupling constants obtained in our simulations. The inset in each figure compares experiment [420] (black curve) and the numerically predicted J-coupling constants at 325K using Sgourakis' [419] (blue curve) and Vuister's [487] (red curve) parameter sets in Karplus equation [225]. Glycine residues are not included because their experimental values are ambiguous.

Table 5.I – Pearson correlation coefficients (PCC) between experimental and calculated J-coupling constants.

	Vuister et al. [487]		Sgourakis et al. [419]		Schmidt et al. [412]	
parameter set	all	part	all	part	all	part
Aβ_{40} – 325K	0.28	0.45	0.29	0.57	0.29	0.57
Aβ_{40} – 300K	0.26	0.40	0.27	0.53	0.27	0.50
Aβ_{42} – 325K	0.43	0.65	0.42	0.72	0.43	0.71
Aβ_{42} – 300K	0.27	0.52	0.29	0.39	0.27	0.44

Glycine residues are not included because their experimental values are ambiguous. The first and the second columns for each parameter set contain, respectively, the PCC obtained when using all the data points and when neglecting the four most problematic residues. For Aβ_{40}, we neglected residues 15, 21, 30 and 40 when using Vuister *et al.* [487], residues 22, 30, 39 and 40 when using Sgourakis *et al.* [419] and residues 22, 30, 39 and 40 when using Schmidt *et al.* [412] parameter sets. For Aβ_{42}, we neglected residues 2, 21, 40 and 41 when using Vuister *et al.*, residues 2, 22, 40 and 41 when using Sgourakis *et al.* and residues 2, 22, 40 and 41 when using Schmidt *et al.* parameter sets.

that obtained by all-atom REMD using OPLS (PCC of 0.66), but is similar to that obtained using all-atom REMD with the AMBER96 force field (PCC of 0.27) [420]. Our moderate correlation for Aβ_{40} is due a shift of more than $\sim$1 Hz for most of the N-terminal residues and four outliers at positions Glu22, Ala30, Val39 and Val40.

Though there is no physical rational for eliminating outliers, to compare with Mitternacht *et al.* [324], we follow their procedure and eliminate the four most significant outliers out of the 24 and 21 experimental data points for Aβ_{40} and Aβ_{42} respectively and compute the PCC using the Schmidt's parameters set [412]. Doing so, our PCCs increase from 0.29 to 0.57 for Aβ_{40} and from 0.43 to 0.71 for Aβ_{42} showing that only a few outlying points strongly impact the PCC (Table 5.I). Using the same approach, our PCC for Aβ_{42}, 0.71, is comparable to the PCC of 0.86 determined by Mitternacht *et al.* [324]

5.4.2 Aβ_{40} properties

Secondary Structure. Averaged over all structures and residues, Aβ_{40} has 7.7 ± 0.1 % β-strand, 3.4 ± 0.1 % α-helix, 50.2 ± 0.1 % turn and 38.8 ± 0.1% random coil. Compared to a recent CD study on Aβ_{40} (8 % α-helix and 24 % β-strand) [355], our ensemble shows a comparable propensity for α-helices, but underestimates the percentage of β-strands. This experiment, however, may overestimate the percentage of β-strands because NMR shows a very weak signal [192, 528] and an earlier CD experiment using a more stringent sample separation protocol reports only $\sim$ 12% of β-strands [243].

The secondary structure propensities for each residue in Aβ_{40} are shown in Figure 5.3. Three β-strands are present from Ala2 to Arg5 with a propensity of 20 to 50%, from Glu10 to His13 and Lys16 to Leu17 with a propensity of 15 to 30 % and from Phe18 to Ala30 with a small propensity ($<$ 1%). The first two β-strands are stabilized by turns at His6-Gly9 and His13-Gln15. The β-strands 2 and 3 are found in competition with two α-helices spanning Glu11-Gln15 with a small probability ($<$ 1%) and Glu22-Asn27 with a probability of up to 25 % as depicted in Figure 5.3. Finally, our analysis also shows a high turn signal ($>$ 60%) between residues 21 and 28 and between residues 33 and 39.

Tertiary Structure. Several contacts play a dominant role in the tertiary structure of Aβ_{40} as shown in Figure 5.4. The C-terminal interacts with the CHC forming hydrophobic contacts between Leu34-Phe19, Val36-Phe19, Met35-Phe19 and Met35-Phe20 with high probabilities

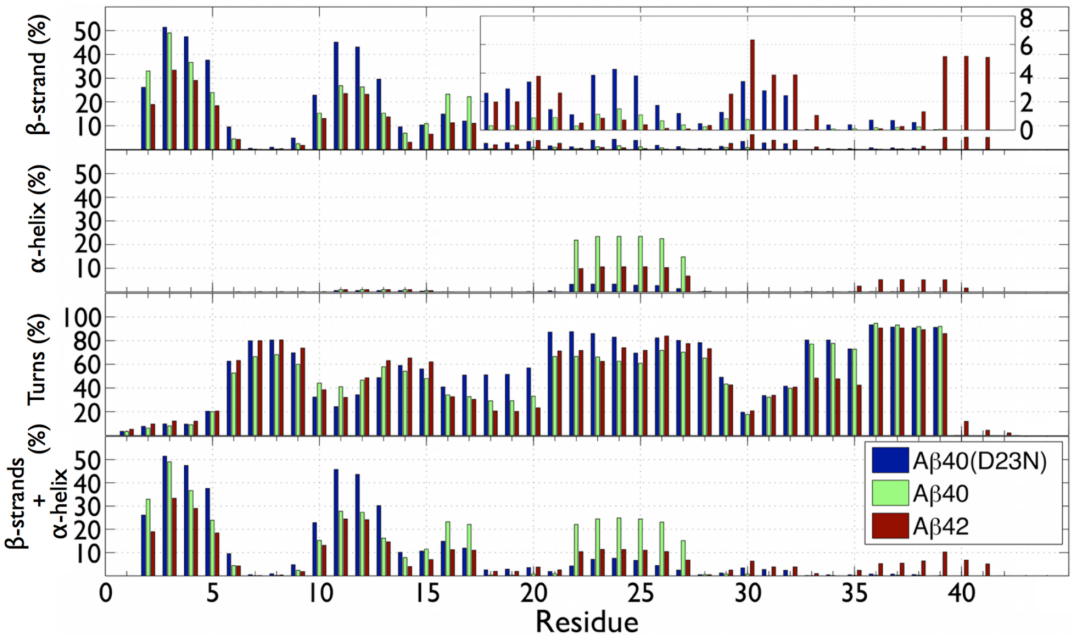


Figure 5.3 – Per residue secondary structure propensities. From top to bottom : β-strand, α-helix, turn and α-helix + β-strand propensities in percentage (%). Aβ_{40} is shown in green, Aβ_{42} in red and Aβ_{40}(D23N) in blue.

and between Ile31-Ile32 and Val18 with slightly lower probabilities. We also observe hydrophobic contacts localized in the CHC, between Leu17-Phe19 and Val18-Phe20, indicating a mostly collapsed core. We also find that the electrostatic Asp23-Lys28 contact, present in about 65% of the morphologies, stabilizes a turn between Ala21 and Lys28.

Column 1 of Figure 5.5 shows the center of the five dominant Aβ_{40} clusters with their respective populations. These five clusters contain 34.3% of all generated conformations. The center of cluster 1 is fully random coil. The centers of clusters 2 to 5 display two β-strands at the N-terminal residues 2-4 and 10-12, leading to a well-formed or a very flexible β-hairpin for the clusters 4-5 and 2-3, respectively. While the cluster centers do not exhibit secondary structure in the CHC or the C-terminal, the loop region is structurally near a α-helical conformation and we find a α-helix at residues 22-26 with $\sim$ 45% probability when averaging over all conformations belonging to the top three clusters. All five clusters display contacts between CHC and the C-terminal. Clusters 1 to 3 also show contacts between residues 1-15 and 21-38, while these contacts are absent in clusters 4 and 5.

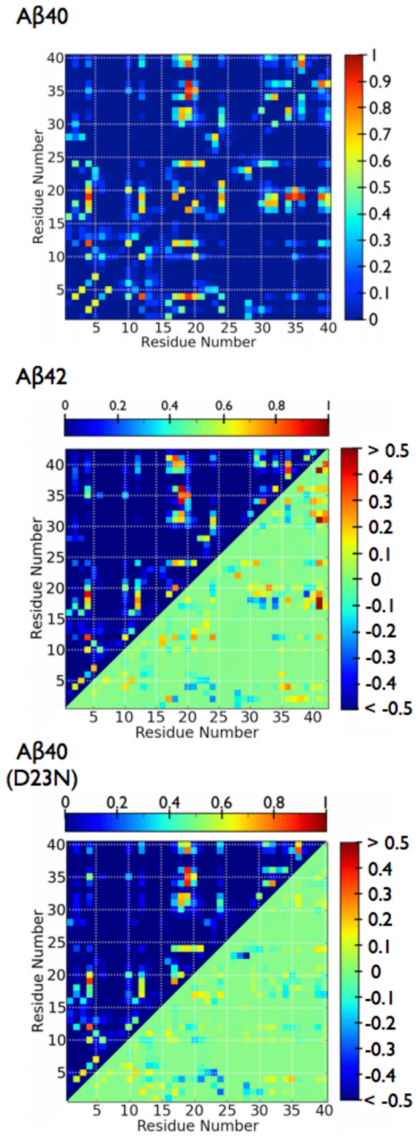


Figure 5.4 – Contact maps between side chains. The contact maps of Aβ_{40}, Aβ_{42} and Aβ_{40}(D23N) are shown in the top, middle and bottom panels respectively. The upper left corner of each panel depicts the propensity of contact between side chains in each alloform. The lower right corner for Aβ_{42} and Aβ_{40}(D23N) displays the contact differences between each alloform and Aβ_{40}. When there is more contacts in Aβ_{40} the propensity is negative (blueish) and the opposite yields a positive propensity (reddish).

5.4.3 Aβ_{42} compared to Aβ_{40}

Secondary Structure. Averaged over all structures and residues, Aβ_{42} shows 2.1 ± 0.1 % α-helix, 6.1 ± 0.1 % β-strand, 48.3 ± 0.1% turn and 43.4 ± 0.1% random coil contents, whereas Aβ_{40} exhibits 3.4 ± 0.1 % α-helix, 7.7 ± 0.1 % β-strand, 50.2 ± 0.1% turn and 38.8 ± 0.1 %

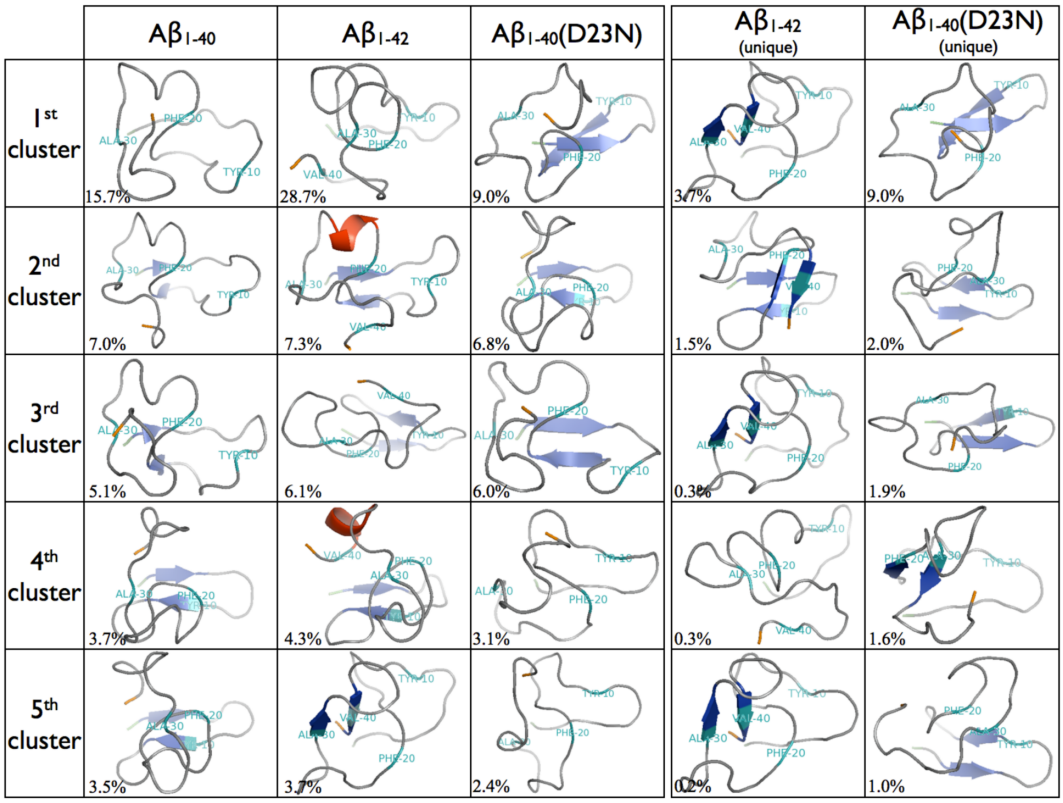


Figure 5.5 – Clusters centers. Centers of the five dominant clusters for Aβ_{40} (first column), Aβ_{42} (second column) and Aβ_{40}(D23N) (third column). The fourth and fifth columns contain respectively the centers of the five dominant clusters of Aβ_{42} and Aβ_{40}(D23N) unique morphologies. The percentages shown are with respect to the total number of structures analyzed. The secondary structure was assigned using STRIDE [164]. The N- and C-termini are shown in green and orange respectively. For all alloforms, Tyr10, Phe20, Ala30 and Val40, for Aβ_{42} only, are shown in teal.

random coil. Thus, these two sequences have the same overall secondary structure composition. It is in the propensity per residue that differences arise.

The secondary structure propensity for each residue is shown in Figure 5.3. As for Aβ_{40}, there are β-strands at the N-terminal at positions Ala2-Arg5, Glu10-His13 and Lys16-Val17, but their propensities are smaller in Aβ_{42}. Noticeably, Aβ_{42} is more prone than Aβ_{40} to form β-strands at the CHC and at positions Ala30-Ile32 and Val39-Ile41 as depicted by the inset in Figure 5.3. Thus, there is a small random coil to β-strand transition at the CHC and the C-terminal in going from Aβ_{40} to Aβ_{42}. Overall, 1.4% of Aβ_{40} and 8.6% of Aβ_{42} structures display β-strands at the CHC or C-terminal. As for Aβ_{40}, there are in Aβ_{42} two α-helices at positions Glu11-Gln15 and Glu22-Asn27 (Figure 5.3). The α-helix at Glu22-Asn27 is however

reduced in going from Aβ_{40} ($\sim$ 22%) to Aβ_{42} ($\sim$ 10%). In addition, we find in Aβ_{42} a third helix at positions Met35-Val40 that is absent in Aβ_{40} (Figure 5.3). On the other hand, both alloform share a very similar turn profile.

Tertiary Structure. Many hydrophobic contacts contribute to the Aβ_{42} monomer morphologies. Most of them are also present in Aβ_{40} as depicted in Figure 5.4 that shows the differences in the contact propensity between the two alloforms. These include a number of contacts between the CHC and the C-terminal, particularly between Met35 and the two residues Phe19 and Phe20. The hydrophobic residue Ile41 increases the overall contact occurrence between the CHC and C-terminal by interacting predominantly with Leu17 and Val18. It also increases contacts inside the C-terminal itself. Finally, the addition of Ile41 and Ala42 causes a small shift of the interactions between the N-terminal and the region 15-35 toward the C-terminal or the N-terminal. This shift may allow the conformational freedom necessary for the CHC to form β-strands and for the region 22-29 to form more easily the specific loop conformation required for fibrillation. Electrostatic interactions also play a role in Aβ_{42} structures. Aβ_{42} and Aβ_{40} have approximately the same propensity of forming a contact between Asp23 and Lys28 ($\sim$ 60%) with a similar contact distance distribution (data not shown).

While the probability of contact occurrence differs between Aβ_{40} and Aβ_{42}, both alloforms have very similar conformational ensemble and only 16 % of all Aβ_{42} conformations are unique, i.e., show a C_α-RMSD greater than 2 Å from any observed Aβ_{40} structures by using residues 1-40. The centers of the five dominant Aβ_{42} clusters are shown in column 2 of Figure 5.5. These five clusters include 50.1% of all conformations. We observe that some of these clusters are structurally similar to the main clusters of Aβ_{40}. For instance, clusters 2 and 4 of Aβ_{42} are respectively similar to clusters 2 and 5 of Aβ_{40}. The third cluster is similar to the cluster 6 of Aβ_{40} that represents 1.2% of its structural ensemble. On the other hand, the first cluster of Aβ_{42} is only similar to marginal clusters of Aβ_{40} ($< 1\%$) and its fifth cluster is not present in Aβ_{40}.

To identify the structural differences between Aβ_{42} and Aβ_{40}, we cluster all unique structures of Aβ_{42}. The first five unique clusters of Aβ_{42} are shown in column 4 of Figure 5.5 and represent 6% of all generated conformations. We observe that these Aβ_{42} unique morphologies display β-strands at the C-terminal and CHC. In particular, the first, third and fifth unique clusters show a β-hairpin motif with strands spanning the C-terminal at residues 30-32 and 39-41. Structurally, these three clusters are near each other differing only at the N-terminal. Another interesting

motif is seen in the second unique cluster where we note that the extremity of the C-terminal may also interact with the CHC forming a β-sheet between residues 18-20 and 39-41.

5.4.4 Aβ_{40}(D23N) compared to Aβ_{40}

Secondary Structure. Averaged over all structures and residues, Aβ_{40}(D23N is composed of $10.1 \pm 0.1\%$ β-strand vs. $7.7 \pm 0.1\%$ in Aβ_{40}, $0.5 \pm 0.1\%$ α-helix vs. $3.4 \pm 0.1\%$ in Aβ_{40}, $56.2 \pm 0.1\%$ turn, which is 6.0% higher than in Aβ_{40}, and $33.2 \pm 0.1\%$ random coil.

Differences are also observed in the per residue secondary structure as shown in Figure 5.3. For example, the β-strand propensity of residues Tyr10-His14 is higher in Aβ_{40}(D23N) than in Aβ_{40} causing a lower β-strand propensity at Gln15-Leu17. Aβ_{40}(D23N) and Aβ_{42} show similar β-strand propensities at positions Val18-Ala21 and Ala30-Ile32 that are absent in Aβ_{40}(Figure 5.3). Overall, 5.0% of Aβ_{40}(D23N) structures vs 1.4% of Aβ_{40} structures display β-strands at the CHC and the C-terminal. From our simulations, we also see that Aβ_{40}(D23N) has a negligible propensity for α-helical configurations, with a small signal at positions 22-27, a feature that may favor a faster appearance of fibril-compatible intermediate oligomers with β-strands at the CHC and the C-terminal in Aβ_{40}(D23N) than in Aβ_{40}.

The turn propensity in the N-terminal, at His6-Gly9 and His13-Gln15, is very similar to Aβ_{40} and Aβ_{42}. Differences are however observed for the CHC, the loop region and the C-terminal. At the CHC, residues 17-20 have a propensity for turn of $\sim$50%, which is higher than in Aβ_{40} and Aβ_{42} by $\sim$20-30% resulting in less random coils. The C-terminal also exhibits a higher propensity for turn in Aβ_{40} and Aβ_{40}(D23N) than in Aβ_{42}, particularly at residues 33-35, generating a longer turn extending from residues 33 to 39. Finally, the loop region has a higher propensity for turn in Aβ_{40}(D23N) than in the wild-type alloforms and the contacts that stabilize this region are very different from Aβ_{40} as described below.

Tertiary Structure. While 16% of Aβ_{42} conformations differ from any observed Aβ_{40} conformations, cluster analysis reveals that 35% of Aβ_{40}(D23N) conformations are distinct from those obtained for Aβ_{40} (C_α-RMSD > 2 Å). Differences in the contact distribution between Aβ_{40} and Aβ_{40}(D23N) are shown in Figure 5.4. For instance, there is a shift of the contacts between the N-terminal and residues 15-30 toward the C-terminal or the N-terminal. While this shift is qualitatively similar to what occurs in Aβ_{42} as described previously, the overall reduction in contacts is more important in Aβ_{40}(D23N). The N-terminal of Aβ_{40}(D23N) looses many

contacts with the region 15-30 without recovering them all with the C-terminal, as shown by Figure 5.4, allowing more conformational freedom. Another main difference is observed for the Asp(Asn)23-Lys28 contact propensity which is reduced by 54% in going from Aβ_{40}(D23N) to Aβ_{40} (Figure 5.4). This induces a rich polymorphism for the region between Ala21 and Ala30 as opposed to the other two alloforms studied. Noticeably, the side chains of Asn23 and Lys28 in Aβ_{40}(D23N) are then unconstrained as they do not form other significant contacts (Figure 5.4).

The centers of the five dominant clusters of Aβ_{40}(D23N) are displayed in the third column of Figure 5.5. These represent 27.3% of all generated conformations. As for Aβ_{40}, they are mostly disordered without any β-strands at the CHC or C-terminal as could be expected since 65% of Aβ_{40}(D23N) and Aβ_{40} conformations are structurally similar (C_α-RMSD < 2 Å). For instance, the centers of the clusters 2 and 3 of Aβ_{40}(D23N) are similar to the centers of the clusters 4 and 1 of Aβ_{40}, respectively. On the other hand, cluster 4 of Aβ_{40}(D23N) is only similar to Aβ_{40} marginal clusters (< 1%) and its first cluster is not found in the structural ensemble of Aβ_{40}.

The Aβ_{40}(D23N) unique clusters are shown in the fifth column of Figure 5.5. The first unique cluster, which has a population of 9%, is also the most dominant cluster for this sequence (see column 3). In term of secondary structure, this cluster is rather unstructured with β-strands only at residues 2-5 and 11-16. The fourth unique cluster displays a β-sheet between residues 18-20 and 30-32 and represents 1.6 % of all visited structures. While the other Aβ_{40}(D23N) unique clusters are as unstructured as Aβ_{40}, their contact distributions differ in two notable ways. (1) Their N-terminals can be isolated from the loop region and interact less with the CHC (unique clusters 1 and 3) ; their contact maps show very few contact between residues 1-15 and 22-29 and between residues 5-15 and 14-21 (data not shown). (2) Or, these unique clusters can have few contacts between residues 1-9 and 22-40 (unique clusters 2 and 5). For these two latter unique clusters, we note also that they exhibit electrostatic interactions between the positively charged Lys28 and the negatively charged Glu11 as Lys28 now interacts very weakly with Asn23 as opposed to WT. Overall, these unique morphologies do not have contacts between the N-terminal and the loop region allowing Aβ_{40}(D23N) to be less compact than Aβ_{40}.

5.5 Discussion

Knowledge of the Aβ conformations at the monomer level in aqueous solution is of utmost importance since the monomer interacts with higher order oligomers [53] and fibrils [148], and

is a building block of the cytotoxic dimer [422]. Both Aβ_{40} and Aβ_{42} monomers have been studied by NMR, CD and computer simulations. Results show that these peptides are described by a distinct ensemble of predominantly random coil structures. While many simulations were performed, [512] very few showed thorough sampling of the relevant morphological ensemble [324, 419]. The similarities and differences of our results with previous experiments and simulations on the Aβ_{40} and Aβ_{42} monomers can be summarized as follows.

Our contact distributions of Aβ_{40} and Aβ_{42} can be compared to the previous all-atom implicit solvent simulation results obtained by Yang and Teplow [512]. We note two differences. First, while we observe a fourth region between residues 30 and 40 in which intra-region contacts are important for both Aβ_{40} and Aβ_{42} ; the role of this region was only identified for Aβ_{42} by Yang and Teplow. Second, there is a slight shift in the contacts between the N-terminal and the loop region. If these regions were seen to interact by contacts between residues 6-10 and 22-28 in the simulations of Yang and Teplow, [512] our results show interactions between residues 1 to 5 and 16 to 28 (Figure 5.4). On the other hand, both simulations agree on several points : the N-terminal interaction with the loop region, the increased number of contacts between the CHC and the C-terminal for Aβ_{42}, the presence of more localized contacts at the C-terminal for Aβ_{42} and the fact that intra-region contacts are more dominant in the four regions corresponding to the turn distribution observed in Figure 5.3.

Inspection of the four sequence regions show interesting features. We find that the N-terminal (residues 1-16) of the the two wild-type alloforms is very similar, with turns at His6-Gly9 and His13-Gln15 stabilizing extended morphologies having β-strands. A weak turn was also observed at Asp7-Glu11 by NMR [192] and two turns were predicted at His6-Gly9 and His14-Lys16 using all-atom REMD simulations [420, 512]. On the other hand, the presence of β-strands at the N-terminal of Aβ_{40} is clearly a matter of debate from experimental and computational studies. While some experiments show that the N-terminal of Aβ_{40} monomer may form β-strands [115], others observe that it is extended and highly fluctuating [277, 523] without any β-strands [192]. Three all-atom REMD simulations report either negligible [420] or low [419, 512] β-strand percentages. In contrast, coarse-grained DMD simulations report a propensity for β-strand of ~ 40 % at Ala2-Phe4 [261]. Taken together, all these studies suggest that the N-terminal may be in rapid exchange between an extended-turn-extended motif free of any H-bonds and β-stranded configurations with a turn at His6-Gly9 being formed most of the time as shown, in our study,

by the propensities in Figure 5.3. We find that Aβ_{42} has also a non-negligible probability to populate a β-hairpin at the N-terminal, a motif that has been predicted recently with two different all-atom potentials [324, 419]. This motif has however a lower probability in Aβ_{42} than in Aβ_{40}. This motif, which was not observed experimentally yet due to its low population, might have consequences on the early formed Aβ_{42} oligomers.

The formation of a loop region between residues 22-29 stabilized by a salt-bridge between Asp23-Lys28 is thought to be one of the rate-limiting steps of Aβ_{40} fibrillation [416]. Recently, the formation of this contact was shown to increase the population of competent fibril-like monomers for Aβ_{10-35} [394]. Here, we observe many morphologies with a turn between Ala21 and Lys28 stabilized by interactions between the side chains of Asp23 and Lys28 in the wild-type sequences (Figures 5.3 and 5.4) with a probability of $\sim 65\%$. This turn was suggested from proteolysis experiment on the Aβ_{40} monomer [265] and its morphology was resolved at the molecular level using solution NMR of the Aβ_{21-30} peptide [265]. The turn ensemble of residues 22-28 for Aβ_{40}, in our simulation, is marginally similar to the two NMR conformations of the Aβ_{21-30} peptide with 20% of our conformations deviating by less than 1.5 Å. These differences can be explain by the presence of contacts between the N-terminal and the loop region with $\sim$ 35% probabilities (Figure 5.4). Overall, our results on Aβ_{40} are consistent with experiments on the Aβ_{40} monomer showing turns at positions 23-29 [277], 20-26 [192] or 24-28 [265]. Our results on Aβ_{40} also agree with previous REMD simulations on the Aβ_{40} monomer [512] and MD simulations on the Aβ_{10-35} monomer and dimer [394] that suggest that the loop region, residues 22-29, needs to undergo structural changes during fibrillation leading to further stabilization of the Asp23-Lys28 contact.

The CHC (residues 17-21) and C-terminal (residues 30-42) are recognized experimentally as the driving regions for aggregation through the formation of β-stranded and extended structures [44, 284, 314]. Experiments revealed that the C-terminal of Aβ_{42} monomer is more rigid than that of Aβ_{40} [511] and displays β-strands at residues Val39-Ile41 [192, 277]. Previous simulations also observed a more structured C-terminal in Aβ_{42} than in Aβ_{40} [261, 419, 420, 512]. In agreement with these experiments and simulations, we observe in our study that the C-terminal residues Ala39-Ile41 modulate, by direct interactions, the appearance of distinct Aβ_{42} morphologies with β-strands at the CHC and C-terminal, that are absent in the Aβ_{40} ensemble. We isolated these morphologies which exhibit a β-sheet between residues 39-41 and 30-32, or bet-

ween residues 39-41 and 18-20 (see fourth column in Figure 5.5). Our results emphasize the important role of Ile41-Ala42 in the monomer morphologies and its role in the early oligomerization process as was observed experimentally [55, 243].

Our findings on the aggregation prone mutant D23N can be summarized as follows. Aβ_{40}(D23N) appears to form a more stable β-hairpin than the two wild-type peptides, as a significant portion of their N-terminal structures. As could be expected, Aβ_{40}(D23N) monomer does not exhibit any significant interaction between Asn23 and Lys28. This region displays rather a high conformational flexibility giving the D23N alloform more freedom to adopt the loop conformations observed in its fibrillar states [471] and therefore reducing the free energy barriers. Interestingly, solid-state NMR indicates that Aβ_{40}(D23N) forms fibrils with multiple morphologies with a majority having anti-parallel β-sheets and a minority having parallel β-sheets [471]. From our simulations, we see that this wide polymorphism may originate in part from the unconstrained side chains of Asn23 and Lys28, the flexibility of the region 21-29 and their overall consequences on Aβ_{40}(D23N) equilibrium ensembles.

Our results show that, while Aβ_{40}(D23N) does not contain Ile41, the mutation D23N increases the β-strand content at the CHC and C-terminal when compared to Aβ_{40} (Figure 5.3). Specifically, one of our unique Aβ_{40}(D23N) morphologies display a β-sheet between residues 18-20 and 30-32 (fifth column in Figure 5.5). Even if its weight of 1.6% is small, this motif is interesting as it involves two region known to be crucial during oligomerization [44, 53, 284].

The higher propensity of Aβ_{40}(D23N) monomer to form β-strands at the CHC and C-terminal and to prefer less collapsed topologies than Aβ_{40} monomer suggests important consequences on its dock-and-lock mechanism when it binds to fibril edge. Previous simulations on the Aβ_{10-40} monomer at the edge of a preformed fibril showed that the fibril edge induce a conversion of the CHC to β-stranded configurations in the monomer [453, 454]. In another computational study, interactions between the C-terminal peptide Met35-Val40 and a fibril were simulated showing a transition of the peptide from random coil to extended configuration upon binding to fibril edge [351]. Taken together, our results suggest that Aβ_{40}(D23N) may exhibit easier conversion to extended configuration upon binding to fibril edge and therefore increase its fibril elongation rate.

Finally, it was observed using photo-induced cross-linking experiments, that D23N promotes the formation of high order oligomers when compared with wild-type Aβ_{40} [53]. High order

oligomers are also formed by Aβ_{42}, which preferably forms pentamer and hexamer, in contrast to Aβ_{40}, which forms low order oligomers such as monomers to tetramers in rapid equilibrium [53]. In our simulation, we observe that both Aβ_{42} and Aβ_{40}(D23N) monomers exhibit a noticeable reduction of contacts between the N-terminal and residues 22-29 and between the N-terminal and residues 18-21 of the CHC (Figure 5.4), a feature that could favor the appearance of high order oligomers by freeing the segment 22-29 and the CHC.

5.6 Conclusion

In this study, we have revisited the structures of the Aβ_{40} and Aβ_{42} monomers using the OPEP force field and the HT-REMD technique, and provided for the first time the structures of the peculiar Aβ_{1-40}(D23N) monomer, which self assemblies without any lag phase to predominant amyloid fibrils with anti-parallel β-sheets. [471]

We observe that even if these three alloforms are mostly disordered at the monomeric level, in agreement with experiments and previous simulations on Aβ_{1-40} and Aβ_{1-42}, striking morphological differences exist. For instance, Aβ_{1-42} have higher β-hairpin propensities at the C-terminal, residues 30-42, than Aβ_{1-40}. Moreover, 16% of all Aβ_{1-42} conformations do not resemble to any Aβ_{1-40} conformations.

The monomeric morphological ensemble of Aβ_{40}(D23N) also present significantly different from that of Aβ_{40} with 35% of unique conformations. In particular, we observe a β-hairpin like motif between residues 18-20 and 30-32 in Aβ_{1-40}(D23N) that is not present in Aβ_{1-40}. Moreover, D23N enhances the the conformational freedom of residues 22-30. It changes the network of contacts, the N-terminal becoming more independent from the rest of the protein, leading to less compact morphologies than the wild-type Aβ_{1-40} peptide. These structural properties could explain why the kinetics and the final amyloid products vary so extensively between the Aβ_{1-40} and Aβ_{40}(D23N) peptides, by increasing the population of the amyloid-competent monomeric state.

5.7 Author contributions

S.C., P.D., and N.M. designed the research ; S.C. performed the research ; S.C. contributed analytic tools ; S.C., P.D., and N.M. analyzed the data ; and S.C., P.D., and N.M. wrote the manuscript.

5.8 Acknowledgement

This work was funded in part by the Canada Research Chairs program, the Fonds québécois de recherche sur la nature et les technologies (FQRNT), the Natural Sciences and Engineering Reasearch Council of Canada (NSERC) and the Fonds de recherche en santé du Québec (FRSQ). Calculations were done on the computers of the Réseau québécois de calcul de haute performance. Financial support of CNRS and the Institut Universitaire de France is also acknowledged.

CHAPITRE 6

DIMÈRES Aβ_{1-40}, Aβ_{1-42} ET Aβ_{1-40}(D23N)

Distinct Dimerization for Various Alloforms of the Amyloid Beta-Protein : Aβ_{1-40}, Aβ_{1-42} and Aβ_{1-40}(D23N)

Sébastien Côté†, Rozita Laghaei†, Philippe Derreumaux‡, and Normand Mousseau†

† Département de Physique and Groupe d'étude des protéines membranaires (GÉPROM), Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec), Canada

‡ Laboratoire de Biochimie Théorique, UPR 9080 CNRS, Institut de Biologie Physico Chimique, Institut Universitaire de France, Université Paris Didero - Paris 7, 13 rue Pierre et Marie Curie, 75005 Paris, France

Article paru : Côté, S., R. Laghaei, P. Derreumaux, and N. Mousseau. 2012. Distinct dimerization for various alloforms of the amyloid beta protein : Aβ_{1-40}, Aβ_{1-42} and Aβ_{1-40}(D23N). *J. Chem. Phys. B*. 116 :4043-4055.

6.1 Abstract

The Amyloid-beta protein is related to Alzheimer's disease and various experiments have shown that oligomers as small as the dimer are cytotoxic. Two alloforms are mainly produced : Aβ_{1-40} and Aβ_{1-42}. They have very different oligomer distributions, and it was recently suggested, from experimental studies, that this variation may originate from structural differences in their dimer structures. Little structural information is available on the Aβ dimer, however, and to complement experimental observations, we simulated the folding of the wild-type Aβ_{1-40} and Aβ_{1-42} dimers as well as the mutated Aβ_{1-40}(D23N) dimer using an accurate coarse-grained force field coupled to Hamiltonian-temperature replica exchange molecular dynamics. The D23N variant impedes the salt-bridge formation between D23-K28 seen in the wild-type Aβ leading to very different fibrillation properties and final amyloid fibrils. Our results show that the Aβ_{1-42} dimer has a higher propensity than the Aβ_{1-40} dimer to form β-strands at the central hydrophobic core (residues 17-21) and at the C-terminal (residues 30-42), which are two segments crucial to the oligomerization of Aβ. The free energy landscape of the Aβ_{1-42} dimer is also broader and more complex than that of the Aβ_{1-40} dimer. Interestingly, D23N also impacts the free energy

landscape by increasing the population of configurations with higher β-strand propensities when compared against Aβ_{40}. In addition, while Aβ_{1-40}(D23N) displays a higher β-strand propensity at the C-terminal, its solvent accessibility does not change with respect to the wild-type sequence. Overall, our results show the strong impact of the two amino acids Ile41-Ala42 and the salt-bridge D23-K28 on the folding of the Aβ dimer.

6.2 Introduction

The hallmark feature of many neurodegenerative diseases such as Parkinson, Huntington, Creutzfeld-Jakob and Alzheimer is the appearance of β-sheet-rich insoluble filamentous deposits in brain tissues [328, 417]. Alzheimer's disease, for instance, is characterized by the formation of extra and intracellular deposits respectively composed of the amyloid β and τ proteins. The amyloid β (Aβ) protein, whose aggregation and oligomer deposition are correlated with the degradation of brain tissues [489], exists in many different alloforms that are produced through the cleavage of the amyloid precursor protein (APP). Aβ_{1-40} and Aβ_{1-42} are the most abundant in neuritic amyloid plaques [198] and the presence of two hydrophobic residues, Ile41 and Ala42, at the C-terminal leads to very distinct oligomer distributions [53, 55, 93] during fibrillation [243, 294, 469] in vitro.

While the exact neurotoxic mechanisms for oligomers are still a matter of debate [406], considerable experimental evidence collected over the last decade shows that metastable Aβ soluble oligomers correlate more with increased neurotoxicity [179]. And while the exact size of these oligomers is not completely clear, even the dimer was recently observed to be synaptotoxic [422]. Both the growth kinetics and toxicity are strongly affected by the exact amino sequence of Aβ peptides. Higher Aβ_{1-42}/Aβ_{1-40} ratio increases toxicity [254]. Aβ_{1-40} and Aβ_{1-42} also show distinct distributions of low order oligomers, which could be due to differences in their dimer equilibrium structures [45]. Mutations can also affect oligomeric growth and the final product. The Iowa familial mutation, Aβ_{1-40}(D23N), for example, fibrillates into anti-parallel β-sheet fibril morphologies without any lag phase [172, 471], contrary to what is observed with both Aβ_{1-40} and Aβ_{1-42}, which show a lag phase and parallel organization [294, 469].

Characterizing the Aβ dimerization at the molecular level is crucial for understanding the origin of the various aggregation properties for these different alloforms [53, 55, 93, 243, 294,

469]. To date, very little experimental information is available for the dimer because it is aggregation-prone and exists in equilibrium with fibrils, monomers and higher-order oligomers [53, 243]. Recently, a combined study using photo-induced cross-linking and circular dichroism (CD) on $A\beta_{1-40}$ showed that the dimerization increases the β-strand propensity and toxicity as compared with the monomer [356].

In the absence of high-resolution structure data such as solution NMR, however, only computer simulations can provide access to detailed structural and kinetic information about the formation of dimers. Until now, the folding of full-length Aβ dimers has been little studied. Monte-Carlo simulations were recently used with an all-atom force field and an implicit solvent to determine the structural differences between the $A\beta_{1-42}$ monomer and dimer, and the impact of the mutants F20E, E22G and E22G/I31E [324, 325]. The conformational differences that are seen involve turns centered in the 20-30 region, hinting at reorganization of this part of the region as a potentially critical step in Aβ aggregation. Discrete molecular dynamics (DMD) simulations with a four-bead coarse-grained potential showed an increase of β-strand propensity during the dimerization of $A\beta_{1-42}$ and $A\beta_{1-40}$ [473, 476]. Shorter Aβ peptides were also studied. A replica-exchange MD (REMD) simulation on the truncated $A\beta_{10-40}$, with the CHARMM force field and implicit solvent, suggests that the dimer is more extended than the monomer due to interstrand contacts [453]. Another REMD simulation using a coarse-grained potential shows that the $A\beta_{16-35}$ monomer and dimer are mostly random coil with low secondary structure signals [88]. The same potential was also used to simulate the C-terminal fragment $A\beta_{29-42}$ [291]. Various simulations were also done on the folding of the full-length Aβ monomer [108, 324, 419, 420, 512].

More simulations on the full-length Aβ dimer including the N-terminal segment (residues 1-16) are needed for three reasons. (1) Mutations [356] or deletions [55] at the N-terminal impact oligomerization and recent experimental and numerical results on an annular morphology of $A\beta_{1-42}$ suggest that this segment might not be unstructured [321] as opposed to previous observations [294]. (2) Unknown structures of the Aβ dimer are cytotoxic [422] and are important building blocks for higher order oligomers [45] ; understanding the structural features of the dimer would help designing more efficient inhibitors [161]. (3) While few folding simulations were done on $A\beta_{1-40}$ [455] and $A\beta_{1-42}$ [325], the only comparison between these sequences was performed using a simple four-bead per residue model with discrete molecular dy-

namics [473]. Moreover, no simulation, to our knowledge, was performed on Aβ_{1-40}(D23N)'s oligomers.

This is why, following our work on the three full-length monomers [108], we choose to investigate the dimerization of Aβ_{1-40}, Aβ_{1-42} and β_{1-40}(D23N) using the coarse-grained OPEP force field, which has been tested on a number of peptides [311], coupled to the Hamiltonian-temperature replica exchange molecular dynamics, for increased sampling efficiency [256].

This manuscript is constructed as follow. After a discussion of the methodology, we analyze the dimer morphologies of Aβ_{1-40}, Aβ_{1-42} and Aβ_{1-40}(D23N). The impact of the addition of the two residues, I41 and A42, and the D23N substitution are discussed with respect to the wild-type Aβ_{1-40}. Finally, we compare our results to previous simulations and experiments, and identify important structural motifs for Aβ dimerization and further aggregation.

6.3 Methods

In this study, we simulated three different alloforms of the amyloid beta protein : Aβ_{1-40}, Aβ_{1-42} and Aβ_{1-40}(D23N) using the OPEP/HT-REMD simulation protocol [256]. The amino acid sequence of Aβ_{1-40} is DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV. For Aβ_{1-42}, two hydrophobic residues, Ile41 and Ala42, are added at the C-terminal. For Aβ_{1-40}(D23N), Asp23 is mutated into Asn23 resulting in a side chain without a net negative charge at neutral pH.

To maximize sampling, temperature replica exchange molecular dynamics, T-REMD [449], is combined with Hamiltonian replica exchange molecular dynamics, H-REMD [166, 256]. We use 22 temperatures : 270, 279, 289, 299, 309, 320, 331, 342, 354, 366, 379, 392, 405, 419, 434, 443, 451, 458, 463, 466, 468 and 470 K. At the highest temperature, we fractionally reduce non-bonded attractive forces using 4 scales : 0.8, 0.6, 0.4 and 0.2. Temperature is controlled using a Berendsen thermostat [40] with a coupling constant of 100 fs, the integration time step is 1.5 fs and exchanges are attempted every 7.5 ps. We save the configurations every 7.5 ps. Bond lengths are constrained with the RATTLE algorithm [9]. All simulations are started from a random extended conformation and each dimer is simulated in a 40 Å-radius sphere with reflecting boundary conditions. Overall, each alloform is simulated for 1250 ns per replica yielding a total simulation time of 32.5 μs per alloform.

Previously, our simulation protocol, HT-REMD coupled with the OPEP force field, has been

tested on the Aβ_{1-40}, Aβ_{1-42} and Aβ_{1-40}(D23N) monomers, leading to good agreement with experiments [108]. For complementarity, we provide a comparison with the experimental values of the chemical shifts measured on the Aβ_{1-40} and Aβ_{1-42} monomers (Figure I.1 and Table I.I).

Force field. To reach longer time-scales, we chose the implicit coarse-grained OPEP potential (Optimized Potential for Efficient peptide structure Prediction) version 3.2 because this force field captures the essential driving forces in protein folding and structure prediction without costing as much computational time as would have an all-atom potential with explicit solvent representation [131, 309–311]. This potential has been shown to recover the native structure of a variety of peptides with widely different secondary and tertiary structures as accurately as all-atom potentials using T-REMD [86] and greedy [309, 310] simulations. Also, OPEP has been applied to study the aggregation of many short and long amyloid peptides such as amylin [256, 257], polyglutamine [255], GNNQQNY [337], KFFE [315] and various segments of Aβ [88, 92, 108, 138, 291, 314, 497]. This potential models each amino acid by C, N, N_H, C_α, O and a single bead (SC) for the side chain. The properties of each side chain are unique and their parameters are finely tuned against thermodynamics and protein structures determined experimentally as described previously [311]. The OPEP force field includes bonded interactions such as bond-lengths, bond angles, improper torsion angles and dihedral angles ; and non-bonded interactions such as Van der Waals and two-body and four-body hydrogen bond potentials.

HT-REMD. Hamiltonian-temperature replica exchange molecular dynamics, HT-REMD [256], is a hybrid of the popular temperature replica exchange molecular dynamics, T-REMD [449], and Hamiltonian replica exchange molecular dynamics, H-REMD [166]. Preliminary tests on the Aβ_{1-42} dimer with the OPEP force field showed that even at very high temperature, the two chains would keep intermolecular interactions, preventing a complete sampling of the conformational space. To reduce this bias, we use a H-REMD scheme at the highest temperature in which we fractionally reduce non-bonded attractive forces, favoring a more complete dissociation of the two chains enhancing conformation space sampling when these replicas move back to lower temperatures. The algorithm has been previously detailed [108, 256].

Analysis. We analyze the secondary and tertiary structures of Aβ_{1-40}, Aβ_{1-42} and Aβ_{1-40}(D23N) at 300K. Secondary structures are predicted using STRIDE [164]. Tertiary structures are analyzed from the network of side chain contacts. Here, a contact is considered formed when two side-chain beads are separated by a distance that is less than the sum of their Van der Waals

radii plus 0.5 Å. To further characterize and extract the relevant morphologies, we employ a two-step clustering procedure. First, the morphologies are regrouped using a RMSD metric with Daura's procedure [119] that finds the biggest cluster, then removes these configurations and repeats iteratively until no structure is left. We choose a relatively selective C_α-RMSD threshold of 2 Å, to provide first screening. Second, all cluster centers are reclustered according to the homology between their contact network using a threshold of 75% homology. In each clustering step, the permutation of the two chains is checked. Molecular graphics images were generated using the PyMOL software (http ://www.pymol.org/). To complement our analysis on the tertiary structure, we also calculated the solvent accessible surface area [146] per residue on all structures reconstructed to all-atom using Scrwl4 [252]. The thermodynamical properties are calculated by WHAM [96]. The free energy landscape is calculated by the histogram analysis method using two reaction coordinates : the percentage of β-strand content and the normalized number of hydrophobic contacts, which is defined as the total number of hydrophobic contacts divided by the number of hydrophobic residues in the alloform. The entropy is extracted using the first law of thermodynamics, $F = E - TS$, and the free energy calculated by WHAM.

Convergence. Convergence of the simulation is assessed in the following way. First, we verify that the entropy as a function of temperature $S(T)$ is time-independent when computed over non overlapping time intervals. This occurs, for the sequences studied here, after 650 ns of simulation at each temperature. Figure 6.1 shows that S(T) remains unchanged for all alloforms when computed on the 650 to 950 ns and 950 to 1250 ns time intervals. Second, computing the cluster distributions, the secondary structures per residue, and the contact maps in both time windows, we confirm that secondary and tertiary structures are well converged. Accordingly, the analysis is performed on the 80 000 structures collected within the 650 to 1250 ns time interval for each alloform.

Naming convention. We focus our analysis on four segments of the peptide that are known experimentally to have a role during the oligomerization [53, 55] : the N-terminal (residues 1-16), the central hydrophobic core (CHC) (residues 17-21), the fibril-loop region (residues 22-28) and the C-terminal (residues 29-40/42). The N-terminal and the fibril-loop region are mostly hydrophilic ; while the CHC and the C-terminal are mostly hydrophobic. In the final amyloid fibril products of the three alloforms, the fibril-loop region forms a loop, and both the CHC and C-terminal form β-sheets ; whereas the N-terminal is mostly disordered [294, 469, 471]. To

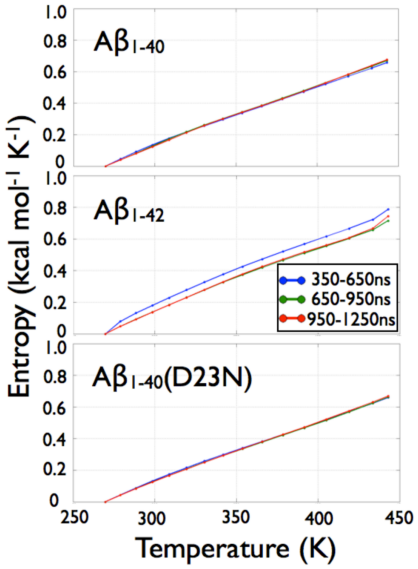


Figure 6.1 – Simulation convergence assessed from the entropy as a function of temperature. From top to bottom, the entropy as a function of temperature for Aβ_{40}, Aβ_{42} and Aβ_{40}(D23N) in the time intervals 350-650 ns, 650-950 ns and 950-1250 ns. The strong similarity between the curves at 650-950 ns and 950-1250 ns for each alloform suggests that our simulations are converged in the time interval 650-1250 ns.

simplify the notation in the Results and Discussion sections, we abbreviate Aβ_{1-40}, Aβ_{1-42} and Aβ_{1-40}(D23N) by Aβ_{40}, Aβ_{42} and Aβ_{40}(D23N) respectively.

6.4 Results

6.4.1 Dimerization of Aβ_{40}

Secondary Structure. The averaged secondary structure propensities over all residues are shown in Table 6.I. We observe that the Aβ_{40} dimer mostly populates turn/random coil with 12.6% of β-strand. The secondary structure per residue reveals that it is not distributed uniformly. The β-strands are mostly localized in three regions : the N-terminal with more than 10% and up to 80% (residues 2-6 and 9-16), and the fibril-loop region and the C-terminal (residues 23-28 and 36-38) with 5-10% ; while the CHC has a weak β-strand propensity that is smaller than 2% (Figure 6.2). α-helices are only populated in the fibril-loop region (5-10%). Finally, residues 7-9, 13-15, 21-28 and 33-39 show a significant propensity for turns, larger than 50%.

Contacts. The networks of total, intrachain and interchain contacts of the Aβ_{40} dimer are shown in Figure 6.3 (first column). We see that the region of highest contact density involves

Table 6.I – Secondary structure averaged over all residues and structures.

	Aβ_{40}		Aβ_{42}		Aβ_{40}(D23N)	
secondary	monomer	dimer	monomer	dimer	monomer	dimer
structure	%	%	%	%	%	%
α-helix	6.0	1.3	4.7	4.4	0.9	0.2
β-strand	11.8	12.6	10.8	30.8	15.9	10.7
Turn	44.8	50.7	44.8	32.4	55.1	52.3
Random coil	37.4	35.4	39.8	32.4	28.0	36.6

The values shown for the dimer are averaged over the converged interval (see Methodology). The monomer results were taken at 300K from our previous study on the full-length Aβ_{40}, Aβ_{42} and Aβ_{40}(D23N) monomers [108]. The statistical errors are small ($< 0.1\%$) according to the interval of confidence on the mean value given by a Bootstrap analysis.

hydrophobic interactions between the CHC (residues 17-20) and the C-terminal (residues 31-36). Other regions with a notable high contact density are : CHC/CHC (residues 17-20/17-20) and C-terminal/C-terminal (residues 36-40/31-36). From the intra and interchain contact maps (panels below), we note that the interactions at the CHC/C-terminal occur both intra and between chains ; while CHC/CHC contacts are almost exclusively interchain. The D23-K28 salt-bridge is present intramolecularly in 55.2% of the configurations ; while it is weakly populated intermolecularly with 4.7% propensity (Table 6.II). Interestingly, there is few contacts, intra or intermolecular, between the N-terminal and the C-terminal. For example, residues 1-16 interact with residues 1-21 ; while they have few contacts with residues 22-40 (fibril-loop region and C-terminal).

Clusters. To complement the analysis on the tertiary structure, we look at the six principal morphologies of the Aβ_{40} dimer totalizing 49% of its recorded structures (first column of Figure 6.4). The secondary structure of the six clusters are shown in Table 6.III. We see that these morphologies display various β-sheets at the N-terminal, and are rather unstructured elsewhere ; although some structural motifs are seen. For instance, the third and fifth clusters have an α-helix at residues 22-26 of the fibril-loop region and the third cluster shows an intramolecular β-sheet in chain 2 between residues 37-38 of the C-terminal and residues 11-13 of the N-terminal. The β-sheets observed in the six dominant clusters are all anti-parallel (Table 6.III).

We also cluster the structures having at least 33% of β-strands at the CHC (2 amino acids) or the C-terminal (3 amino acids) in either chain since these marginal morphologies may be more

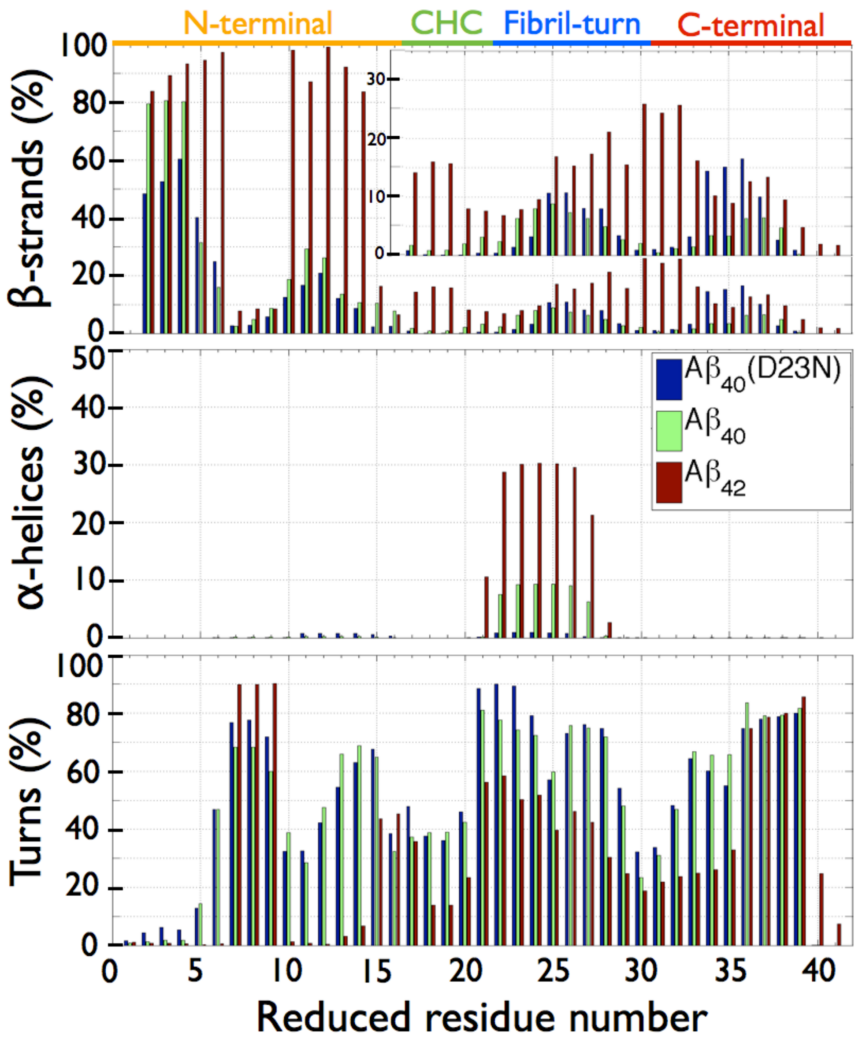


Figure 6.2 – Per residue secondary structure. From top to bottom, β-strands, α-helices and turns propensities for Aβ_{40} (green), Aβ_{42} (red) and Aβ_{40}(D23N) (blue).

prone to oligomerize [284]. These structures represent 14% of the total ensemble and the first six clusters are shown in Figure 6.5. We see different types of motifs involving the C-terminal : it forms a two-stranded β-sheet with the C-terminal (interchain in cluster 4), or with the fibril-loop region (intrachain in clusters 2 and 6), or with the N-terminal (intrachain in cluster 1 or interchain in cluster 3). On the other hand, the CHC is not involved in any β-sheet with the C-terminal for these clusters ; it instead forms a β-sheet with the fibril-loop region (interchain in cluster 5). Only the third cluster shows parallel β-sheets (between the N- and C-terminals) ; while the other clusters show anti-parallel β-sheets (Table 6.III).

Free energy landscape. The free energy landscape, which is plotted as a function of the

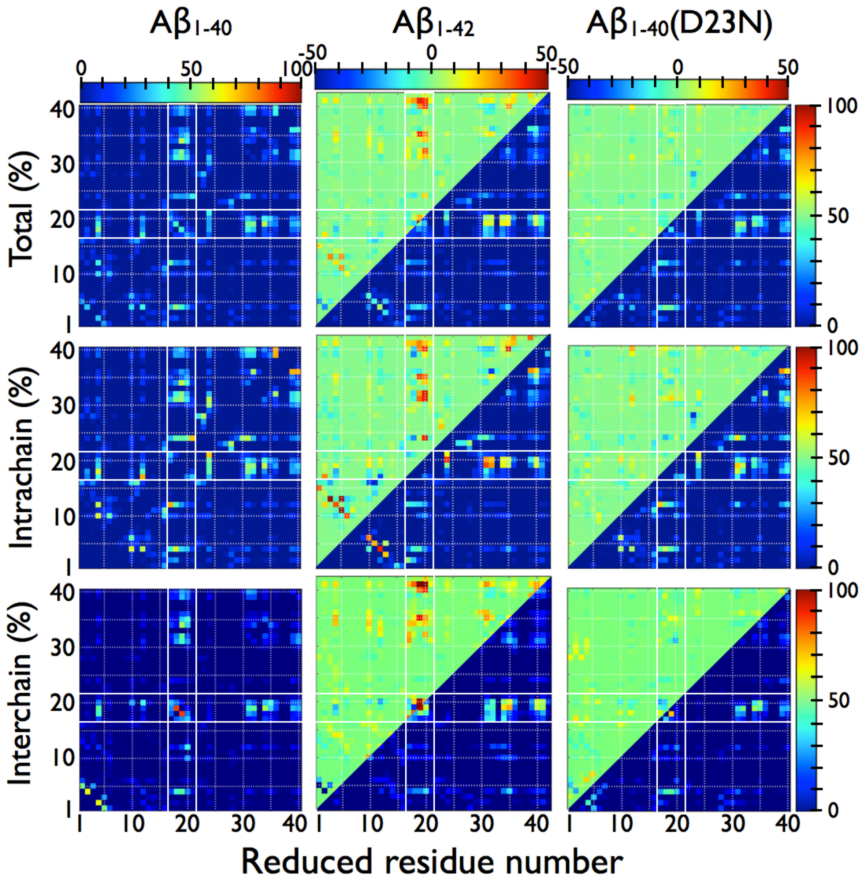


Figure 6.3 – Network of side-chain contacts. The percentages of total (first row), intrachain (second row) and interchain (third row) contacts are depicted. The total contact propensity is calculated from % intra plus % inter. The first column shows the values for the Aβ_{40} dimer and the bottom-right corner maps of the second and third columns show that values for the Aβ_{42} and Aβ_{40}(D23N) dimers, respectively. The dimers of Aβ_{42} (second column) and Aβ_{40}(D23N) (third column) are compared to the dimer of Aβ_{40} in the top-left corner maps. The percentage difference is calculated from the % in the Aβ_{42} or Aβ_{40}(D23N) dimers minus the % in the Aβ_{40} dimer. The white lines, which are found between residues 16 and 17, and between residues 21 and 22, delimit the CHC region.

percentage of β-strand content (x-axis) and the normalized number of hydrophobic contacts (y-axis), is shown in Figure 6.6 (first row). We observe that most morphologies are found in a region between 4.5-5.25 normalized hydrophobic contacts and between 12-22% β-strand propensity. There, we observe the three deepest free energy minima, which are centered around 4.8 normalized hydrophobic contacts and 13%, 15% and 20% β-strand, and which are separated by weak free energy barriers (1.2 and 0.8 kcal/mol respectively). Morphologies with a β-strand propensity higher than 25% are sparsely populated.

Table 6.II – Propensities of K28-D23 and K28-E22 contacts.

	Aβ_{40}		Aβ_{42}		Aβ_{40}(D23N)	
	intra	inter	intra	inter	intra	inter
contact	%	%	%	%	%	%
K28 - E22	35	4	27	0	29	3
K28 - D(N)23	55	5	54	0	14	1
K28 - other D or E	8	4	4	6	7	11
K28 - E22	+ 12		- 17		- 6	
K28 - D(N)23	- 18		- 18		- 1	
K28 - other D or E	- 12		- 9		- 4	

Rows 1 to 3 show the propensities of intra and intermolecular contacts between K28 and D(N)23 or E22 in the Aβ_{40}, Aβ_{42} and Aβ_{40}(D23N) dimers. Rows 4 to 6 show the difference in propensities between the dimer and the monomer (% dimer minus % monomer). The values shown for the dimer are extracted from Figure 6.2. The monomer results were taken at 300K from our previous study on the Aβ_{40}, Aβ_{42} and Aβ_{40}(D23N) monomers [108]. The third and sixth rows show the mean contact propensities between K28 and D1, E3, D7 or E11, which are the other negatively charged residues in Aβ. The statistical errors are small $< 1\%$ according to the interval of confidence on the mean value given by a Bootstrap analysis.

Table 6.III – Secondary structure of Aβ_{40} dominant morphologies.

	clusters in Figure 6.4		clusters in Figure 6.5	
cluster no.	res. no. (chain no.)	motif	res. no. (chain no.)	motif
1	8-12 (#1) / 2-6 (#2)	a.-p. β	22-27 (#1)	α
			2-4 (#1) / 2-4 (#2)	a.-p. β
			10-12 (#2) / 36-38 (#2)	a.-p. β
2	2-7 (#1) / 11-16 (#1) / 2-4 (#2)	a.-p. β	2-4 (#1) / 10-12 (#1)	a.-p. β
			25-28 (#1) / 34-37 (#1)	a.-p. β
3	22-26 (#1)	α	2-4 (#1) / 2-4 (#2)	a.p. β
	2-4 (#1) / 2-4 (#2)	a.-p. β	10-12 (#1) / 36-38 (#2)	p. β
	11-13 (#2) / 37-38 (#2)	a.-p. β	37-38 (#1) / 15-16 (#2)	p. β
4	2-4 (#1) / 2-5 (#2) / 11-13 (#2)	a.-p. β	9-12 (#1) / 2-5 (#2)	a.-p. β
			34-36 (#1) / 34-36 (#2)	a.-p. β
5	22-26 (#1)	α	2-4 (#1) / 2-4 (#2)	a.-p. β
	2-4 (#1) / 2-11 (#2) / 23-27 (#2)	a.-p. β	23-30 (#1) / 20-28 (#2)	a.-p. β
6	2-4 (#1) / 2-4 (#2) / 14-15 (#2)	a.-p. β	2-4 (#1) / 10-12 (#1)	a.-p. β
			25-28 (#1) / 34-37 (#1)	a.-p. β

The clusters of Figure 6.4 represent the dominant morphologies, while the clusters of Figure 6.5 represent the dominant morphologies having at least 33% β-strand content at the CHC or at the C-terminal. The motifs indicated in the Table are α-helix (α), and anti-parallel (a.-p.) and parallel (p.) β-sheets (β).

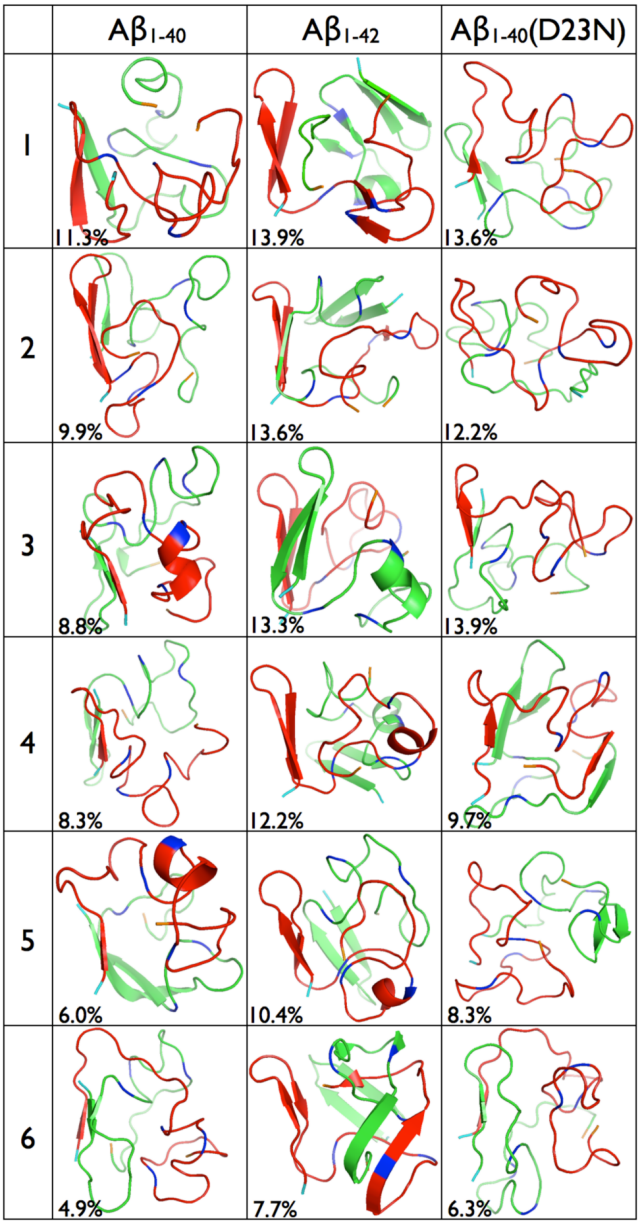


Figure 6.4 – Dominant morphologies. The six main clusters' centers are shown for Aβ_{40} (first column), Aβ_{42} (second column) and Aβ_{40}(D23N) (third column). The N-terminal and the C-terminal are shown in teal and orange respectively. The first residue in each of the peptide region is shown in blue : Leu17 (CHC), Glu22 (fibril-loop region) and Ala30 (C-terminal). The first chain is shown in red, the second in green.

6.4.2 Dimerization of Aβ_{42}

Secondary Structure. The Aβ_{42} dimer has a much higher β-strand propensity (31%) than the Aβ_{40} dimer (13%) as shown in Table 6.I. This gain leads to an equivalent reduction in turn

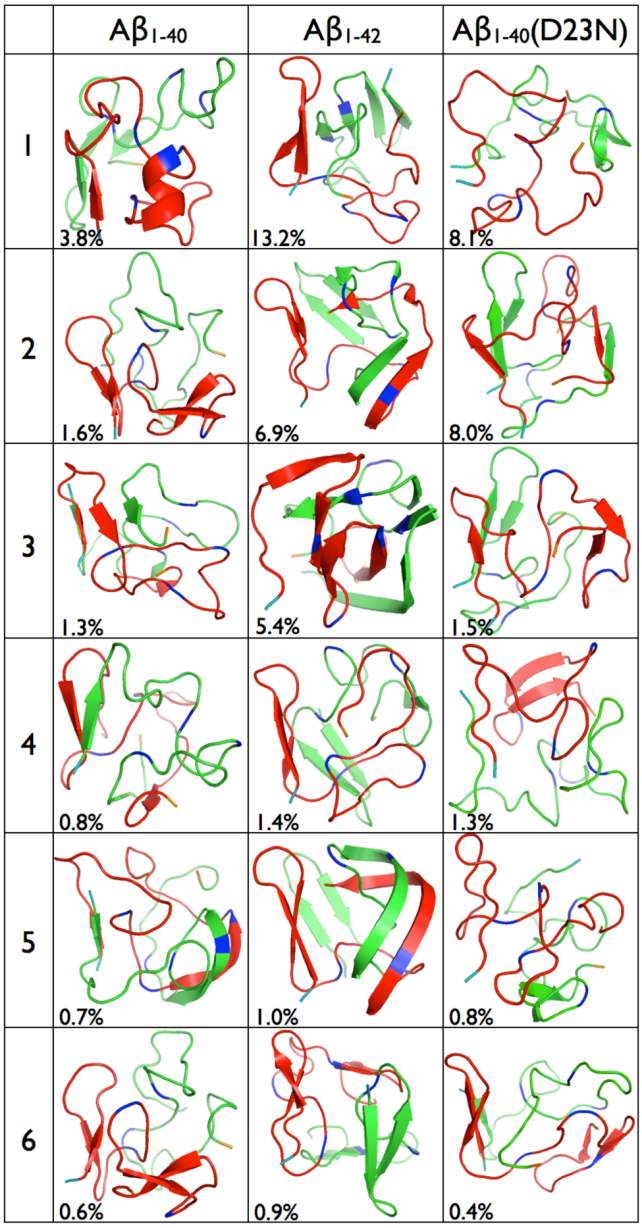


Figure 6.5 – Dominant morphologies having β-strands at the CHC or C-terminal. The centers of the six main clusters having β-strands for more than 33 % of the residues at the CHC or at the C-terminal in either chain are shown for Aβ_{40} (first column), Aβ_{42} (second column) and Aβ_{40}(D23N) (third column). The N-terminal and the C-terminal are shown in teal and orange respectively. The first residue in each of the peptide regions is shown in blue : Leu17 (CHC), Glu22 (fibril-loop region) and Ala30 (C-terminal). The first chain is shown in red, the second in green.

propensity that falls to 32%. This leaves 32% random coil and 4% α-helix, which are similar to Aβ_{40} propensities. The Aβ_{42} dimer enhances the formation of β-strands at the expense of turns

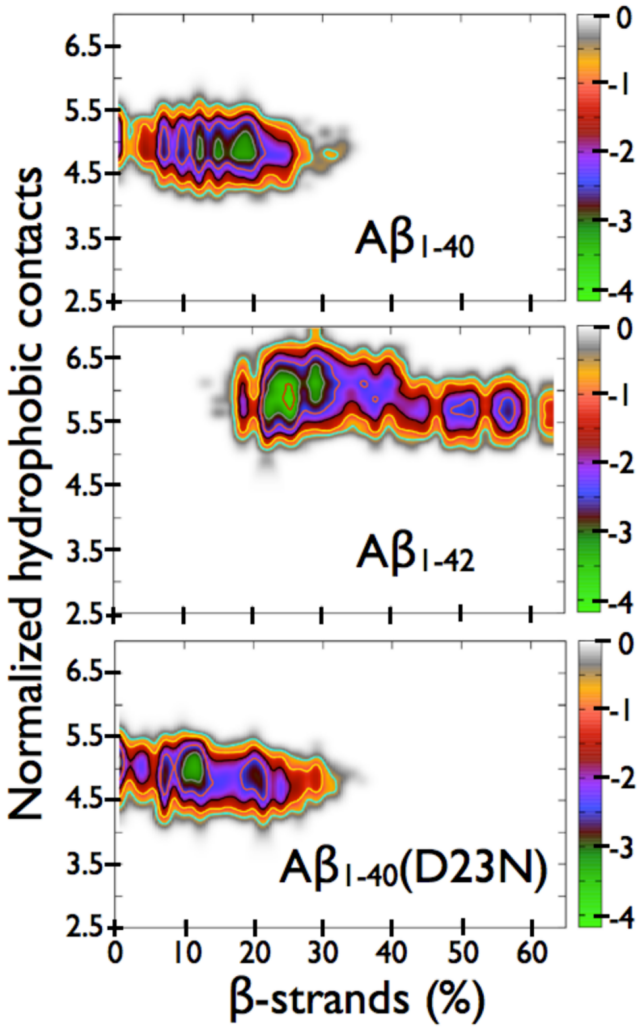


Figure 6.6 – Free energy profile. From top to bottom, the free energy profile in kcal/mol of Aβ_{40}, Aβ_{42} and Aβ_{40}(D23N) projected onto the β-strand content (x-axis) and the normalized number of hydrophobic contacts (y-axis). The normalized number of hydrophobic contacts is calculated from the total number of hydrophobic contacts divided by the number of hydrophobic residues in the alloform.

for almost all residues when compared to the Aβ_{40} dimer (Figure 6.2). Residues 2-6 and 10-14 of the N-terminal populate β-strands with 80% probability or more. The β-strand propensity is 7-15% at residues 17-21 (CHC) and 10-25% at residues 25-38 (fibril-loop region and C-terminal), which is higher than Aβ_{40} as this alloforms shows $<$ 2% (CHC) and 5-10% (C-terminal). The turn propensity of Aβ_{42} is lower than Aβ_{40} for all residues ; except residues 7-9. For example, the turn at residues 13-15, which is present in Aβ_{40}, is absent from Aβ_{42}. The resulting turn distribution with a propensity greater than 50% is thus more localized to residues 7-9, 21-24

and 36-39 (Figure 6.2). On the other hand, as for Aβ_{40}, the Aβ_{42} dimer exhibits a polymorphic fibril-loop region displaying α-helix, β-strand, turn and random coil in competition.

Contacts. Contact maps for Aβ_{42} are provided in Figure 6.3 (second column). Similarly to Aβ_{40}, the highest contact density region is localized at the CHC/C-terminal region (residues 17-20/31-41). The other notable regions of high contact density are : CHC/CHC (residues 17-20/17-20) and C-terminal/C-terminal (residues 31-36/34-41) (top row). The main difference resides in that this alloform has more contacts overall than the Aβ_{40} dimer. More precisely, hydrophobic contact propensities between the following regions are strengthened by the addition of Ile41 and Ala42 : CHC/CHC (interchain), CHC/C-terminal (intra and interchain) and C-terminal/C-terminal (intra and interchain) as shown in the intrachain and interchain contact maps (middle and bottom rows respectively). As for the charged residue K28, it displays a slightly reduced propensity for the formation of an intramolecular contact with D23 and a much lower propensity to interact with E22 in the Aβ_{42} dimer than in Aβ_{40} (Table 6.II). Intermolecular contact propensities between K28 and D23 or E22 are very low in Aβ_{42} as for Aβ_{40}.

Clusters. As expected from the secondary structure analysis, the clusters of the Aβ_{42} dimer are more ordered than for the Aβ_{40} dimer (see the second column of Figure 6.4). The positions of the secondary structure elements in each cluster are shown in Table 6.IV. The six clusters represent 71% of the 80 000 structures recorded for Aβ_{42} indicating that, already at the dimer level, Aβ_{42} cannot be described as fully random coil/turn configurations. In contrast with Aβ_{40}, the N-terminal of these morphologies has a high propensity to populate a β-hairpin motif, and the CHC can be involved in a β-sheet as seen in the first cluster. In the second cluster, the N-terminal of each chain forms a three-stranded β-sheet with the fibril-loop region of the other chain. In the third cluster, the β-hairpin motif at the N-terminal of each chain interact forming a quasi four-stranded β-sheet, and an α-helix is formed in the fibril-loop region. The fourth and fifth clusters have an intramolecular two-stranded β-sheet at their N-terminals with an α-helix in the fibril-loop region. Finally, the sixth cluster displays an almost completely extended C-terminal which forms a β-sheet with the fibril-loop region of the other chain. In terms of orientation, the β-sheets observed in these six clusters are all anti-parallel (Table 6.IV).

We also look at the morphologies having at least 33% of β-strand at the CHC (2 amino acids) or C-terminal (4 amino acids) in either chains (second column of Figure 6.5), a criterion that is met by more than 38% of Aβ_{42} visited morphologies, compared with 14% for Aβ_{40}. Among

Table 6.IV – Secondary structure of Aβ_{42} dominant morphologies.

	clusters in Figure 6.4		clusters in Figure 6.5	
cluster no.	res. no. (chain no.)	motif	res. no. (chain no.)	motif
1	22-27 (#2)	α	2-6 (#1) / 12-14 (#1)	a.-p. β
	2-6 (#1) / 10-14 (#1)	a.-p. β	2-6 (#2) / 10-14 (#2)	a.-p. β
	19-23 (#1) / 28-30 (#1)	a.-p. β	17-19 (#2) / 30-32 (#2)	a.-p. β
	2-6 (#2) / 10-14 (#2)	a.-p. β		
	17-19 (#2) / 30-32 (#2)	a.-p. β		
2	2-6 (#1) / 10-14 (#1) / 25-28 (#2)	a.-p. β	4-6 (#1) / 10-12 (#1)	a.-p. β
	23-24 (#1) / 2-6 (#2) / 10-14 (#2)	a.-p. β	27-33 (#1) / 33-39 (#2)	a.-p. β
			36-37 (#1) / 29-30 (#2)	a.-p. β
			40-41 (#1) / 25-26 (#2)	a.-p. β
			2-6 (#2) / 10-14 (#2)	a.-p. β
3	22-27 (#2)	α	7-13 (#1) / 30-33 (#2)	a.-p. β
	2-6 (#1) / 10-14 (#1)	a.-p. β	15-19 (#1) / 24-28 (#1)	a.-p. β
	2-6 (#2) / 10-15 (#2)	a.-p. β	30-37 (#1) / 3-13 (#2)	a.-p. β
			15-18 (#2) / 25-28 (#2)	a.-p. β
4	23-27 (#1)	α	2-6 (#1) / 12-14 (#1)	a.-p. β
	21-27 (#2)	α	2-6 (#2) / 10-14 (#2)	a.-p. β
	2-6 (#1) / 10-14 (#1)	a.-p. β	32-33 (#2) / 37-38 (#2)	a.-p. β
	2-6 (#2) / 10-14 (#2)	a.-p. β		
5	22-26 (#1)	α	2-6 (#1) / 10-14 (#1)	a.-p. β
	2-6 (#1) / 12-14 (#1)	a.-p. β	27-41 (#1) / 25-39 (#2)	a.-p. β
	2-6 (#2) / 10-14 (#2)	a.-p. β	3-6 (#2) / 10-13 (#2)	a.-p. β
6	4-6 (#1) / 10-12 (#1)	a.-.p β	4-7 (#1) / 10-12 (#1)	a.-p. β
	27-33 (#1) / 33-39 (#2)	a.-.p β	30-32 (#1) / 36-38 (#1)	a.-p. β
	36-37 (#1) / 29-30 (#2)	a.-.p β	2-6 (#2) / 10-14 (#2)	a.-p. β
	40-41 (#1) / 25-26 (#2)	a.-.p β	30-32 (#2) / 36-38 (#2)	a.-p. β
	2-6 (#2) / 10-14 (#2)	a.-.p β		

Same as Table 6.III.

the important features of this subgroup of conformations, we note an important role for the C-terminal which forms either inter or intrachain β-sheets with itself (respectively cluster 2 and clusters 4 and 6), as well as with the CHC (intrachain in cluster 1), with the fibril-loop region (interchain in clusters 2 and 5), and with the N-terminal (interchain in cluster 3). Interestingly, the β-sheet motif between the CHC and C-terminal is not observed for Aβ_{40}. Moreover, the β-sheets at the C-terminal are longer than what is observed for Aβ_{40} and we see for instance a long anti-parallel intermolecular β-sheet between the two C-terminals in the fifth cluster. This latter motif is interesting as one side of this β-stranded C-terminal is completely exposed to the solvent and it might be possible that such motif promotes, during further oligomerization, the formation

of an intermolecular β-sheet with the C-terminal of a third chain. In these six clusters, all β-sheets have an anti-parallel orientation (Table 6.IV). Finally, we note that no α-helix is observed in these clusters for Aβ_{42} as opposed to Aβ_{40}.

Free energy landscape. The free energy profile for the Aβ_{42} dimer is shown in Figure 6.6 (second row). It shows that most morphologies have between 5.25-6.5 hydrophobic contact per hydrophobic residues and 15-60% β-strand propensity. As a result, the phase space visited is shifted toward more hydrophobic contacts and higher β-strand contents than in Aβ_{40}, in agreement with our analysis of their secondary structures, contact propensities and clusters. In details, the two deepest free energy minima, which are separated by a $\sim$ 0.9 kcal/mol energy barrier, are found in a region between 5.25-6.5 normalized hydrophobic contacts and 22-32% β-strands. Other important free energy minima, which are separated by higher energy barriers ($\sim$ 1.5 kcal/mol), are found at higher β-strand contents, namely at 35-40%, 48-53% and 55-59%. A shallower minimum is found at even higher β-strand content (62-63%). The addition of Ile41 and Ala42 in Aβ_{42} energetically favors morphologies with higher hydrophobic contact propensity per hydrophobic residue ($>$ 5.5) and β-strand propensity ($>$ 30%) than in Aβ_{40}. Overall, the free energy landscape of the Aβ_{1-42} dimer is broader and more complex than that of the Aβ_{1-40} dimer.

Are these morphologies stable for the Aβ_{40} dimer ? We previously showed that the Aβ_{42} and Aβ_{40} dimers visit very different morphologies. To ensure that this difference is real and not associated with sampling limitations, we truncated the last two C-terminal residues, Ile41 and Ala42, from the 22+4 replicas of the Aβ_{42} dimer HT-REMD in the middle of the converge interval, at t=925 ns, and relaunched the simulation to ascertain their stability. After only 25 ns, we observe that the secondary structure of A$\beta_{42}\Delta$(41-42) relaxes toward the equilibrated data for Aβ_{40} at 300K, the temperature of analysis. To confirm that this is not only a brief structural reorganization, we continued the simulation. After 325 ns, the β-strand, α-helix and turn distributions have relaxed to the values obtained for Aβ_{40} (Figure I.2). The same observation is made at 270K, the lowest temperature of analysis. The difference observed between the various alloforms is therefore not due to a sampling artifact.

6.4.3 Dimerization of Aβ_{40}(D23N)

Secondary Structure. The secondary structure of the Aβ_{40}(D23N) dimer is shown in Figure 6.2. While the averaged propensities over all residues for Aβ_{40}(D23N) is similar to Aβ_{40} (Table 6.I), the per residue propensities differ notably. Namely, the β-strand propensity of Aβ_{40}(D23N) is lower at the N-terminal, but higher at residues 34-37, where we observe a lower turn propensity, leading to a more extended C-terminal when compared to WT. The α-helix propensity is also much lower in the fibril-loop region enhancing turn (residues 20-24) and β-strand (residues 25-28) propensities. Overall, the secondary structure distribution of Aβ_{40}(D23N) can be described by four regions where β-strands are present with propensities greater than 5% (residues 2-6, 9-14, 25-28 and 34-37) and four regions where turns are found with propensities greater than 50% (residues 7-9, 13-15, 21-29 and 33-39). We note, in particular, that residues 17-21 of the CHC are almost free of any β-strand and α-helix propensities.

Contacts. The region with the highest contact density for Aβ_{40}(D23N) is the CHC/C-terminal (residues 17-20/31-36), as for Aβ_{40} (Figure 6.3, third row). The CHC/CHC (residues 17-20/17-20) and C-terminal/C-terminal (residues 36-40/31-36) regions also display high contact density. The mutation D23N causes an overall small increase in the number of inter and intramolecular contacts, particularly between the CHC/C-terminal regions and the CHC/N-terminal regions. This increase is, however, far less important than what is observed when Aβ_{42} is compared to Aβ_{40}. In spite of this general rise in contacts, some specific regions show a decrease when compared to WT. This is the case for the interchain contacts between the CHC/CHC regions (residues 17-18/20), as well as for K28, which less frequently forms a contact with D23 that has been mutated to an asparagine (Table 6.II). Even if freed from this salt bridge, K28 does not interact significantly more with the other negatively charged residues.

Clusters. The first six clusters of Aβ_{40}(D23N), which represent 64% of its recorded structures, are relatively bared of secondary structure (Figure 6.4 and Table 6.V). The secondary structure elements of its clusters are shown in Table 6.V. The first cluster only displays a three-stranded β-sheet between residues 2-3 (chain 1), 2-5 (chain 2) and 10-12 (chain 2) at the N-terminal without any other secondary structure. The second cluster is completely disordered, while the third cluster shows only a short β-sheet between residues 2-4 (chain 1) and 2-4 (chain 2). A different motif from the WT main morphologies is observed, however, in the fourth cluster that shows an intermolecular β-sheet at the C-terminal between residues 33-36 (chain 1) and

34-37 (chain 2) as well as a three-stranded β-sheet at the N-terminal (residues 4-6 of chain 1, and residues 2-6 and 12-14 of chain 2). The C-terminal residues 34-37 of the fifth cluster are also involved in an intramolecular β-sheet with the fibril-loop region (residues 25-28). Finally, the sixth cluster is mostly disordered, as the third cluster, having only a short β-sheet at the N-terminal between residues 2-4 (chain 1) and 4-6 (chain 2). As opposed to the clusters of the other alloforms, some of Aβ_{40}(D23N)'s clusters display a parallel β-sheet at the N-terminal (clusters 1, 4 and 6) as shown in Table 6.V. In contrast, the β-sheets involving other parts of the peptide are anti-parallel.

Table 6.V – Secondary structure of Aβ_{40}(D23N) dominant morphologies.

	clusters in Figure 6.4		clusters in Figure 6.5	
cluster no.	res. no. (chain no.)	motif	res. no. (chain no.)	motif
1	2-3 (#1) / 2-5 (#2) / 10-12 (#2)	a.-p./p. β	25-28 (#2) / 34-37 (#2)	a.-p. β
2	-	-	4-7 (#1) / 2-6 (#2) / 12-14 (#2)	a.-p./p. β
			33-36 (#1) / 34-37 (#2)	a.-p. β
3	2-4 (#1) / 2-4 (#2)	a.-p. β	4-6 (#1) / 2-6 (#2) / 12-14 (#2)	a.-p./p. β
			25-27 (#1) / 35-37 (#1)	a.-p. β
4	4-6 (#1) / 2-6 (#2) / 12-14 (#2)	a.-p./p. β	25-29 (#1) / 34-38 (#1)	a.-p. β
	33-36 (#1) / 34-37 (#2)	a.-p. β		
5	25-28 (#2) / 34-37 (#2)	a.-p. β	25-28 (#2) / 34-37 (#2)	a.-p. β
6	2-4 (#1) / 4-6 (#2)	p. β	2-6 (#1) / 10-13 (#1) / 2-4 (#2)	a.-p./p. β
			28-30 (#1) / 37-39 (#1)	a.-p. β

Same as Table 6.III.

The clusters formed from the configurations containing at least 33% of β-strand at the CHC (2 amino acids) or at the C-terminal (3 amino acids) in either chains represent 22% of Aβ_{40}(D23N) structures, which is more than Aβ_{40} (14%). The first six clusters are shown in Figure 6.5. We observe that these remain mostly unstructured, similarly to Aβ_{40}, but with a few original motifs. For example, the first and fifth clusters display a β-hairpin motif between the fibril-loop region (residues 25-28 of chain 2) and C-terminal (residues 34-37 of chain 2) whereas all the other residues are unstructured. A similar motif is observed in the fourth cluster between residues 25-29 and 34-38 in chain 1. In the second cluster, an intermolecular anti-parallel β-sheet between the C-terminal is seen (residues 33-36 of chain 1 and residues 34-37 of chain 2). In contrast, similarly to Aβ_{40} (cluster 6), the C-terminal of the third and sixth clusters interacts with the fibril-loop region as well as having a small β-sheet at the N-terminal. As for the dominant morphologies, these clusters only show anti-parallel β-sheets outside the N-terminal ;

while some parallel organization is observed at the N-terminal (Table 6.V).

Free energy landscape. The free energy profile of Aβ_{40}(D23N), which is shown in Figure 6.6, is more similar to Aβ_{40} profile than Aβ_{42}. As Aβ_{40}, most morphologies populated are found between 4.25-5.5 hydrophobic contacts per hydrophobic residues and 0-30% β-strand propensity. However, the minima layout exhibits striking differences. There are two free energy basins (around 6-15% and 18-22% β-strand content) that are separated by a free energy barrier of $\sim$ 1.1 kcal/mol. Also, the population of mostly disordered morphologies ($<$ 5% β-strand) is larger. In addition, D23N enhances the population of Aβ morphologies with 25-35% β-strand propensities. On the other hand, the number of hydrophobic contacts per hydrophobic residues is mostly unaffected.

6.5 Discussion

Direct experimental atomic-level information on the dimer of the Aβ protein is not available because trapping a specific transient oligomer is extremely challenging in aqueous solution [53, 243, 356]. Although computational studies can fill this gap, very few simulations were done on the aggregation of the full-length Aβ dimer [325, 455, 473] and, to our knowledge, only one simulation compares the two wild-type sequences, Aβ_{1-40} and Aβ_{1-42}. [473]. Moreover, to our knowledge, no simulation has been performed on the Aβ_{1-40}(D23N) oligomers. Characterizing the Aβ dimer is important as it is the smallest oligomer that can serve as building block for higher order toxic oligomers [45] and it has itself been recognized neurotoxic [422]. Here, following our characterization of the three isolated full-length monomers [108], we simulate the physiologically relevant Aβ_{1-40}, Aβ_{1-42} and Aβ_{1-40}(D23N) dimers using the OPEP force field [311] coupled to HT-REMD [256]. Our results reveal notable differences and similarities between these alloforms, which we relate to previous studies on Aβ, and they can be summarized as follow.

6.5.1 Dimerization

We first look at the structural changes occurring during Aβ dimerization by comparing with our previous results on the monomer [108] at 300K using the same protocol and force field. We expect more β-strands in the dimer than the monomer from CD experiments [356], as well as T-REMD [453] and DMD [473] simulations on Aβ_{40}. In agreement with these works, our results

for Aβ_{40} show a slight increase in the averaged total β-strand propensity upon dimerization (Table 6.I). This increase is particularly localized at the fibril-loop region and C-terminal, while we observe a destabilization at the N-terminal (residues 10-16) as shown in Figure 6.7. Results are similar but of larger magnitude for the Aβ_{40}(D23N) dimer : the β-strand propensity is also reduced at the N-terminal and increased at the C-terminal suggesting a shift of β-strands from the N-terminal to the C-terminal during dimerization for both Aβ_{40} and Aβ_{40}(D23N). It is in Aβ_{42}, however, that the difference is the largest as this alloform exhibits a significant increase of β-strand propensity in all regions leading to a more significant organization at this level. Overall, we observe an important role for the C-terminal during the dimerization, as it adopts a more extended conformation that favors greatly the formation of β-strands (Figure 6.7).

We also observe that Aβ dimerization for these three alloforms is characterized by a global increase in hydrophobic contacts particularly between the CHC/CHC, CHC/C-terminal and C-terminal/C-terminal regions (Figure 6.8). Moreover, the solvent accessible surface area of the CHC and C-terminal residues for these three alloforms is significantly reduced in the dimer (panel B of Figure 6.9). These observations confirm that the hydrophobic residues of the Aβ monomer are not properly buried, as expected from experiments showing that the Aβ monomer is mostly random coil in solution [192, 277, 511, 528]. The dimer formation is then largely driven by the large energy gain associated with the burial of these hydrophobic residues as confirmed by the formation of a large number of intermolecular hydrophobic contacts, and the appearance of more intramolecular contacts for some residues (Figure 6.8).

Taken together these results show that intermolecular contacts due to dimerization lead to a more extended peptide particularly at the C-terminal of Aβ_{40}, as has been previously seen in a T-REMD simulation [453]. Moreover, we observe that this is generalizable to other alloforms, at least for Aβ_{42} and Aβ_{40}(D23N), and that it might be a common feature of Aβ dimerization. Interestingly, each of these three alloforms exhibits strong intermolecular contacts at the CHC (Figure 6.8). This common feature could be targeted by chemical compounds interacting with the CHC region to forbid, already during dimerization, the formation of intermolecular hydrophobic contacts there. It would be interesting to see if other alloforms, especially those having a mutation in this region such as Aβ_{42}(F19P) [44], also share this feature.

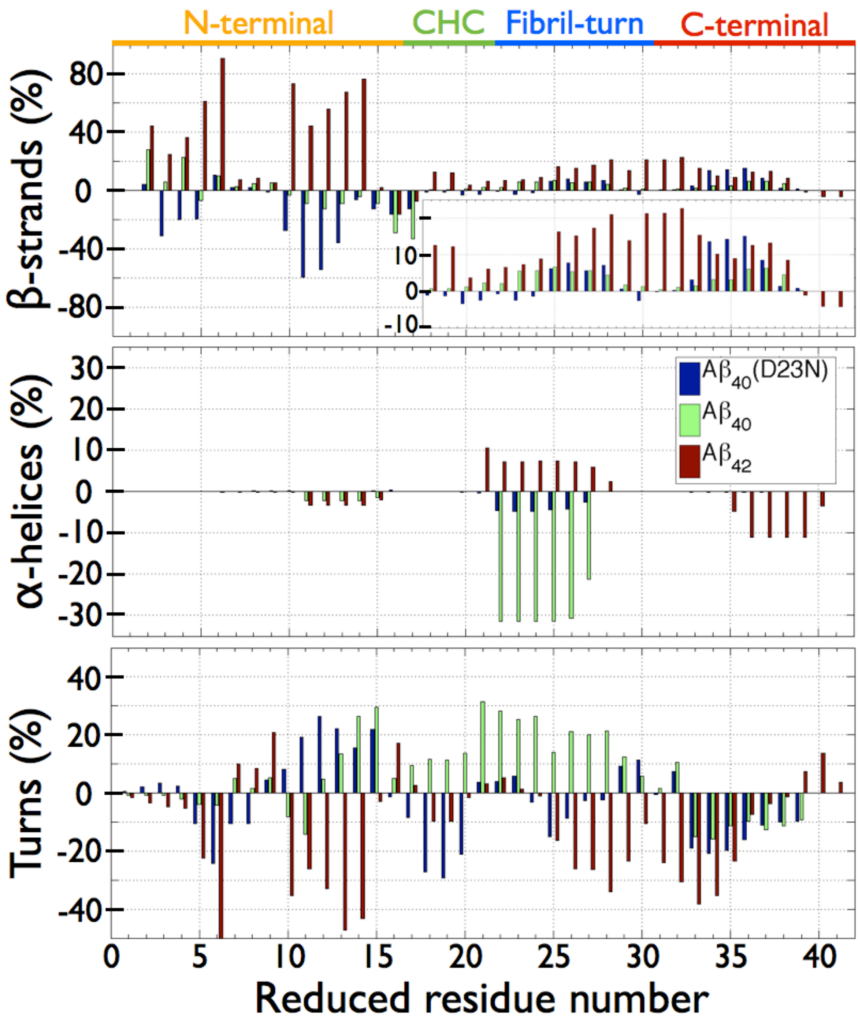


Figure 6.7 – Per residue secondary structure difference between the dimer and the monomer. From top to bottom, β-strands, α-helices and turns propensities differences for Aβ_{40} (green), Aβ_{42} (red) and Aβ_{40}(D23N) (blue). The percentages shown represent : % in the dimer minus % in the monomer.

6.5.2 Dimer morphologies

Second, we compare the structural features of the Aβ_{42} and Aβ_{40}(D23N) dimers to the Aβ_{40} dimer. Below, we summarize our findings by focusing on the four regions that play a determinant role in Aβ aggregation [55] : N-terminal (residues 1-16) [321, 356], CHC (residues 17-21) [44, 94, 284], fibril-loop (residues 22-29) [416] and C-terminal (residues 30-42) [284].

The N-terminal is unstructured in both Aβ_{40} and Aβ_{42} fibril morphologies [294, 469], but a recent study combining experiments and simulations shows otherwise for an annular, transient morphology of Aβ_{42} [321]. In this morphology, these authors observed that the N-terminal is

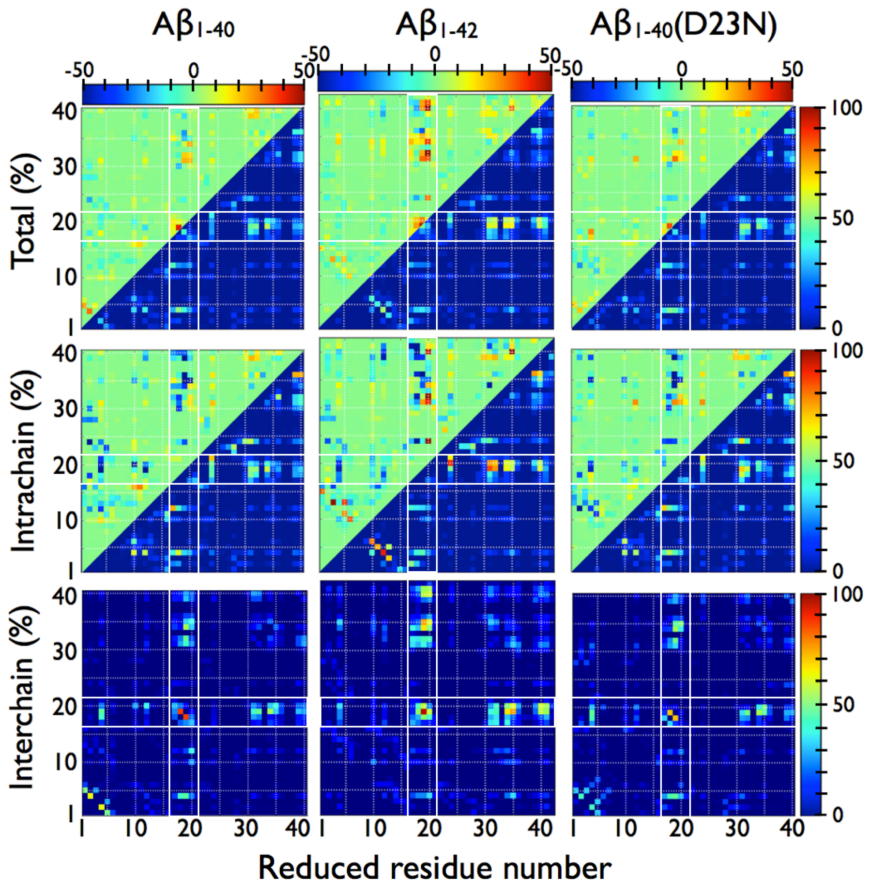


Figure 6.8 – Side chain contact network of the dimer compared to the monomer. The contact maps' lower right corner display the contact propensity of total (first row), intrachain (second row) and interchain (third row) contacts for Aβ_{40} (first column), Aβ_{42} (second column) and Aβ_{40}(D23N) (third column) ; while the contact maps' upper left corner show the impact of dimerization on the contact propensities from the % in the dimer minus the % in the monomer. The total contact propensity is calculated from % intra plus % inter. The white lines, which are found between residues 16 and 17, and between residues 21 and 22, delimit the CHC region.

found to form intermolecular β-sheets with other N-terminals and it interacts with the fibril-loop region suggesting an important role for the N-terminal segment of Aβ, which is nonetheless discarded in most simulations. For the monomer, some simulations showed that Aβ_{42} might form β-hairpin motifs at the N-terminal [108, 419] or form β-sheets with other parts of the peptide [473, 512].

Our results show a complex role for the N-terminal in the dimer. While the N-terminal of Aβ_{42} adopts very often a β-hairpin motif, the N-terminal of Aβ_{40} and Aβ_{40}(D23N) populates mostly random coil configurations or small intra and intermolecular β-sheets as shown in Fi-

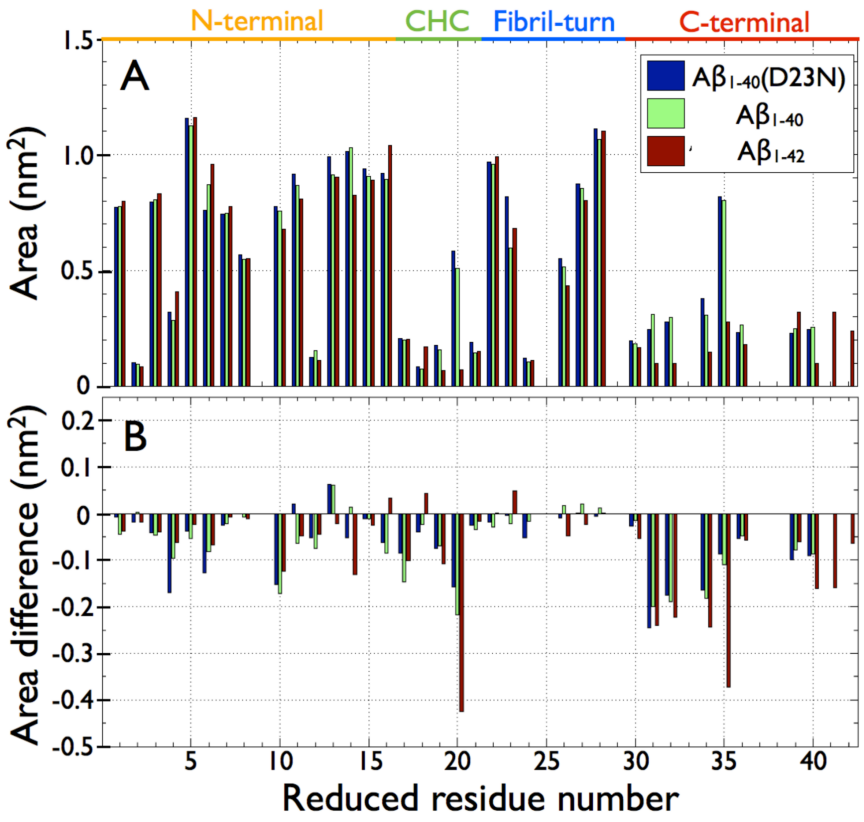


Figure 6.9 – Per residue solvent accessible surface area. (A) In the top panel, the per residue solvent accessible surface area for the Aβ_{40} (green), Aβ_{42} (red) and Aβ_{40}(D23N) (blue) dimers are compared. (B) The surface accessible area difference between the dimer and the monomer are shown in the lower panel. Surface area differences are obtained from the dimer % minus the monomer %.

gures 6.2 and 6.4, in agreement with a previous DMD simulation on the Aβ_{40} dimer [473]. Compared to this same study, it seems that OPEP overestimates the β-sheet propensity at the N-terminal of Aβ_{42} [473]. However, a recent MC simulation showed otherwise as the N-terminal of Aβ_{42} populates a β-hairpin centered at residues 14-15 with a probability of 74-82% ; while we found a hairpin centered at residues 7-9. No experimental results on the structural ensemble of the Aβ_{42} dimer is available, to our knowledge, that could settle this matter.

The N-terminal is localized at the surface of the protein and it creates an interface between the solvent and the hydrophobic residues of the CHC and C-terminal (Figures 6.4 and 6.5) reducing their solvent accessibility. The solvent accessible surface area (SASA) per residue including the N-terminal is shown in Figure 6.9. We observe that the hydrophilic N-terminal and fibril-loop region are mainly exposed to the solvent ; except residues Ala2, Val12 and Val24 in all alloforms. While the hydrophobic CHC and C-terminal regions are essentially buried, their

SASA is lower in Aβ_{42} than Aβ_{40} ; the variant D23N, on the other hand, is very similar to wild-type Aβ_{40}. To quantify the screening done by the N-terminal on these regions, we re-calculated the SASA per residue without considering the N-terminal (residues 1-16). The results clearly show that the N-terminal is responsible in part for shielding the CHC and the C-terminal, as well as the charged residues D23 and K28 in each alloform (Figure I.3). This shielding is least important for Aβ_{42}, suggesting a greater intrinsic stability for its hydrophobic core. To unveil the origin of the lower SASA of the CHC and C-terminal of the Aβ_{42} dimer, we re-calculated again the SASA without considering Ile41 and Ala42, which are the only difference between Aβ_{42} and Aβ_{40}. Surprisingly, their direct impact on SASA is not more important than the N-terminal (Figure I.4) ; instead their effect on the tertiary structure is caused indirectly by an overall increase of hydrophobic contact propensities (Figure 6.3).

Taken together our results on the N-terminal show that this region is important to shield the CHC and C-terminal hydrophobic residues from the solvent. In the Aβ_{42} dimer, its role seems secondary compared to the impact of Ile41-Ala42 on the tertiary structure folding in agreement with a previous experiment showing that Aβ_{42} early oligomerization is weakly impacted by the truncation of the N-terminal residues 1-10, as opposed to Aβ_{40}, and that it is strongly impacted by the truncation of Ile41 and Ala42, A$\beta_{42}\Delta$(41-42) [55].

The fibril-loop region has been suggested to be important during nucleation as a preformed salt-bridge between D23-K28 was shown to greatly enhance the nucleation and fibrillation rates of Aβ_{1-40} [416]. Focusing on the Aβ_{21-30} fragment, experiments show the presence of a turn between residues 24-28 that populates two main conformations with a salt-bridge between K28 and D23 or K28 and E22 [265]. These salt-bridge populations could have been overestimated [27], however, and while computational studies on the full-length Aβ monomer also showed the presence of these two salt-bridges, the 21-30 segment was found to be rather polymorphic suggesting that further stabilization of this contact is needed during oligomerization [108, 512].

Here, we observe that the fibril-loop region remains polymorphic in the dimer, visiting α-helix, β-strand, turn and random coil conformations. The D23-K28 contact is still populated in the WT dimer, but its propensity is reduced compared to the monomer (Table 6.II). This contact may be easily solvated, like in the monomer, since both D23 and K28 are still highly accessible to the solvent (Figure 6.9). This observation has to be contrasted against a previous simulation on the Aβ_{10-35} in which it was observed that dimerization slightly reduces the D23-K28 salt-bridge

solvent accessibility [394]. But, these authors also observed that the salt-bridge can still be easily solvated without a Lactam bridge construct between D23 and K28. Taken together their results and ours suggest that further stabilization of the D23-K28 salt-bridge and the fibril-loop occur at a later stage of assembly. For Aβ_{40}(D23N), as expected, the neutralized side-chain of residue 23 has a significantly smaller contact propensity with K28, which become less constrained and does not show a significantly higher contact propensity with the other negatively charged residues in Aβ (Table 6.II). Surprisingly, this additional freedom due to a single point mutation is sufficient to stabilize the formation of β-strands at the C-terminal as discussed below.

Both the CHC and the C-terminal are known to play an important role during oligomerization [44, 45, 55] and fibrillation [294, 321, 469, 471] through the formation of extended conformations and β-sheets [284]. Previous experiments and simulations have shown that the C-terminal of the Aβ_{42} monomer is more extended than Aβ_{40} [108, 277, 473, 511, 512]. Similar observations were also made for the CHC [108, 192]. These regions were also shown to be important to modulate the tertiary structure of the Aβ_{42} monomer [324, 419].

Here, our results suggest that these two regions in the dimer are significantly more extended with higher β-strand and lower turn propensities in Aβ_{42} than in Aβ_{40} (Figure 6.2). (1) The C-terminal of Aβ_{42} is involved in more diverse β-stranded motifs with higher propensities than in Aβ_{40} (Figures 6.2, 6.4 and 6.5, and Tables 6.III and 6.IV). For example, the second and fifth clusters of Figure 6.5 exhibit a β-sheet at the C-terminal that is laterally accessible to the solvent on one side. This motif, which is not observed for Aβ_{40}, could promote Aβ_{42} early oligomerization and be important during nucleation by favoring the recruitment of a third chain at the C-terminal level to elongate the β-sheet in a morphology similar to the fibril morphology : parallel β-sheet made of the C-terminal of each Aβ peptide [294, 321]. More investigation will be needed to clarify the role of this motif in the assembly of higher order oligomers. (2) The CHC of Aβ_{42} is involved in β-stranded motifs with the C-terminal (first cluster in Figure 6.4) as opposed to Aβ_{40} that is almost unstructured there (Figure 6.2). This region is also involved in significantly more contacts in Aβ_{42} (Figure 6.3). The CHC thus appears more important in Aβ_{42} than in Aβ_{40} and we expect that mutations in this region would have a greater impact on Aβ_{42}. Interestingly, Bitan *et al.* observed experimentally that mutations at the CHC strongly disturb Aβ_{42} early oligomerization whereas Aβ_{40} is only weakly affected [55]. Moreover, these authors showed that the opposite is observed when the negatively charged Glu22 or Asp23 is mutated.

These observations support our results, which show that these two residues are involved in fewer interactions with the positively charged K28 in Aβ_{42} than in Aβ_{40} (Table 6.II). The C-terminal residues Ile41 and Ala42 are thus critical to the Aβ_{42} dimer morphologies by enhancing the overall β-strand and hydrophobic contact propensities, and decreasing the electrostatic contact propensities.

In our simulations, the Aβ_{40}(D23N) dimer exhibits structural motifs that differ from those observed in both Aβ_{40} and Aβ_{42}. While its C-terminal has a higher β-strand propensity than in Aβ_{40}, its CHC is almost devoid of any secondary structure as opposed to Aβ_{42}. Its main morphologies are rather unstructured, but some exhibit interesting structural motifs that are weakly populated in the Aβ_{40} ensemble with, for example, an intermolecular anti-parallel β-sheet at the C-terminal (fourth cluster in Figure 6.4 and other less populated clusters). This motif, which appears already in the dimer, could be important for the nucleation of Aβ_{40}(D23N) as some of its fibril morphologies are characterized by anti-parallel, instead of the common parallel, β-sheet at the C-terminal [471].

6.5.3 Comparison to previous studies

Our results on the full-length Aβ_{40} and Aβ_{42} dimers can be compared to the other three published folding simulations on these alloforms [325, 455, 473].

Our results on the Aβ_{40} dimer can be compared to the results of Takeda *et al.* [453, 455] obtained using the CHARMM19 force field with SASA implicit solvent model coupled to REMD. In term of secondary structure, Takeda *et al.* observed 6% β-strand, 19% α-helix and 49% turn for their Aβ_{10-40} structural ensemble [453] and they later observed that the structural ensemble is weakly impacted by the addition of residues 1-9 [455] ; whereas we found higher β-strand (12.6%) and very low α-helix (1.3%) propensities. Turn propensity is however similar (them : 49% vs. us : 44.8%). Their contact network is also different with most of the intermolecular contacts being between the N-terminal and the rest of the sequence ; whereas we observe high intermolecular contact densities between the CHC and the C-terminal, and between the CHC of the two chains (their Figure 5 vs. our Figure 6.2).

Our Aβ_{42} dimer simulation can be compared to the recent work of Mitternacht *et al.* [325] that used an all-atom force field with implicit solvent representation coupled to Monte Carlo. In terms of secondary structure, these authors observed that the Aβ_{42} dimer marginally populates

α-helices, and mostly adopts anti-parallel β-sheet configurations, in agreement with our observations (their text vs. our Figure 6.2, and Tables 6.I and 6.IV). Similarly to us, they observed that the turn distribution is localized at residues 7-10 (us : 7-9), 14-15 (us : no), 20-21 and 25-26 (us : 21-24), and 35-38 (us : 36-39), where we consider a turn when its propensity is at least 50%. In term of intramolecular contacts, their results show that the C-terminal/C-terminal and CHC/C-terminal regions have high contact densities, agreeing with us ; however their region of highest contact density, between CHC and residues 9-12, is absent in our simulation (their Figure 3 vs. our Figure 6.3). Their interchain contacts are also similar to ours with the highest densities being between the C-terminal/C-terminal, the CHC/CHC and the C-terminal/CHC regions. Their C-terminal/CHC intermolecular contact propensities are, however, less populated then ours. In their simulation, the contact D23-K28 is populated $\sim$30-40% of the time, which is less than in our simulation (53.5%). Finally, their six dominant clusters display similar β-sheet interactions to our clusters : C-terminal/C-terminal (intra), N-terminal/fibril-loop (intra), CHC/C-terminal (intra) and C-terminal/fibril-loop (intra) (their Figure 6 vs. our Figures 6.5 and 6.4). We observe more intermolecular β-sheet motifs though and our β-sheets are shorter in most cases.

The impact of adding two residues to Aβ_{40} was also studied by Urbanc *et al.* [473] with a four-bead coarse-grained force field with implicit solvent representation coupled to DMD simulations. In terms of secondary structure, their results show higher β-strand propensities at the CHC and at the C-terminal in Aβ_{42} than in Aβ_{40} (their Figure 6 vs. our Figure 6.2), in agreement with our results. Their turn distributions for Aβ_{40} and Aβ_{42} are similar to ours except at residues 36-39 for which we observe similar turn propensity for these two alloforms ($\sim$80%) ; whereas they observe lower propensity for Aβ_{40} ($\sim$40%) (their Figure 7 vs. our Figure 6.2). They also found a marginal population of α-helices, in agreement with our results. In terms of contact network, our results agree on the following regions of high contact density for both alloforms : CHC/CHC (intermolecular), CHC/C-terminal (inter and intramolecular) and C-terminal/C-terminal (intramolecular) (their Figures S6 and S7 vs. our Figure 6.3). On the other hand, our results show smaller contact propensity involving the N-terminal (intra and intermolecularly) and the C-terminal/C-terminal (intermolecularly). Finally, in terms of solvent accessible surface area, they observed that the Aβ_{42} and Aβ_{40} dimers have almost the same solvent accessibility per residue as opposed to our results showing that Aβ_{42} residues at the CHC and at the C-terminal are more buried than in Aβ_{40} (their Figure S5 vs. our Figure 6.9).

Overall, we see that some of our results on Aβ_{40} and Aβ_{42} agree with previous computational studies ; however there are variabilities concerning the net propensities of contacts, secondary structures and solvent accessible area that are associated with the various force fields and simulation conditions used. Ultimately, new experimental studies will help to validate these simulations.

6.5.4 Role of the dimer on oligomerization

In light of our results, we now analyze the role that the Aβ dimer could have in further oligomerization. Recent experiments showed that the early oligomerization of Aβ_{40} and Aβ_{42} differ markedly due to structural differences already present at the dimer and tetramer levels [45]. Our simulations show that the Aβ_{42} dimer ensemble is energetically more favorable, with more hydrophobic contacts, less solvent accessible surface area for the CHC and C-terminal, and more secondary structure motifs than the Aβ_{40} dimer ensemble. This suggests that, at later stages of assembly, Aβ_{40} will need to form more intermolecular hydrophobic contacts to reduce their solvent accessible surface area. We can then expect the Aβ_{40} tetramer to be dominated by collapsed and rather amorphous conformations. Our results for the Aβ_{42} dimer show, on the contrary, that it is dominated by configurations with few unfavorable solvent accessible hydrophobic residues suggesting that the Aβ_{42} tetramer could easily adopt more open and structured configurations. This agrees with what has been observed and hypothesized in the aforementioned experiment [45], namely the Aβ_{42} tetramer is more open ; Aβ_{40} is more globular. The globular aspect of the Aβ_{40} tetramer was also observed from recent folding simulations [241, 473]. Fewer numerical results exists for Aβ_{42} tetramer and folding has only been studied using DMD simulation which shows, contrary to the previously mentioned experimental results, that the Aβ_{42} oligomers populate globular configurations [473]. Other molecular dynamic simulations on the folding of Aβ_{40} and Aβ_{42} using the same simulation protocol as benchmark and sampling enhancing techniques such as HT-REMD will be needed to evaluate more carefully the role of the dimer in the formation of the tetramer morphologies as well as the structural features characterizing them.

Finally, it is known experimentally that both D23N and Aβ_{42} favor the appearance of higher-order oligomers [55] with an increased fibrillation rate as compared to Aβ_{40} [294, 471]. Our results suggest that the origin of this behaviour is distinct for Aβ_{42} and Aβ_{40}(D23N) as they

exhibit very different structural motifs already at the dimer level ; even if we had previously identified similarities between their monomers [108]. Formation of β-strands at the CHC is crucial to the oligomerization of Aβ, but the D23N alloform, as discussed above, does not have β-strand at the CHC, as opposed to Aβ_{42}, hinting that this motif might be important later on during Aβ_{40}(D23N) oligomerization/fibrillation. The fibril-loop region of both alloforms has also different contact and secondary structure propensities (Table 6.II and Figure 6.2, respectively). Moreover, Aβ_{42} displays significantly more hydrophobic interactions than Aβ_{40}(D23N) (Figure 6.3). These differences lead us to think that diverse and independent structural features may favor the oligomerization and fibrillation of Aβ. The different mechanisms of Aβ assembly – formation of extended structures at the CHC and C-terminal through β-sheet [284], a salt-bridge between D23-K28 [416], a loop between residues 22-29 and the strand-loop-strand motif [294, 469, 471] – would thus happen at different stages of aggregation and have distinct predominance for different alloforms.

6.6 Conclusion

In this study, we analyzed in details the impact of dimerization on the secondary and tertiary structures of three physiologically relevant Aβ alloforms : Aβ_{1-40}, Aβ_{1-42} and Aβ_{1-40}(D23N). Our study combines secondary structure and contact propensities, free energy landscapes, solvent accessible surface and clustering. Our simulations clarifies the role of Ile41 and Ala42, and D23N on the dimerization of the Aβ peptide.

For Aβ_{1-42}, the two additional hydrophobic residues at the C-terminal, Ile41 and Ala42, have a strong impact on folding. They increase the overall hydrophobic contact and β-strand propensities, and reduce the presence of electrostatic contacts as well as the solvent accessibility of the residues at the CHC and C-terminal. The C-terminal of Aβ_{1-42} is also involved in more diverse interactions with the other parts of the peptide. Morphologies with globally higher β-strand and hydrophobic contact propensities are favored as shown by the free energy analysis.

The mutation D23N enhances the conformational freedom of the positively charged K28 causing, surprisingly, an increase of β-strand propensity at the C-terminal relative to the wild-type Aβ_{1-40}. While its predominant morphologies are rather unstructured with no secondary structure at the CHC, the presence of β-stranded motifs at the C-terminal such as an intermolecular anti-parallel β-sheet, which population is marginal in Aβ_{1-40} morphological ensemble,

could be important to facilitate nucleation. In addition, the free energy landscape of the D23N variant shows that this alloform increases the population of configurations with larger β-strand propensities relative to the wild-type Aβ_{1-40}.

Overall, our results reveal that oligomers as small as the dimer of these three alloforms already have very different structural motifs suggesting different oligomerization pathways in agreement with previous experiments on Aβ. They also emphasize that more simulations on higher order oligomers, using enhancing sampling algorithms such as HT-REMD, will be needed to further understand the role of the dimer in later stage of assemblies and how the oligomerization is impacted by the distinct structural features of different Aβ alloforms.

6.7 Supporting Material

In Annexe I, table showing the Pearson correlation coefficients comparing the monomers chemical shifts to experiment and figures showing monomers chemical shifts per residue, time evolution of the secondary structure of A$\beta_{1-42}\Delta$(41–42), and SASA without the N-terminal residues or Ile41–Ala42.

6.8 Author contributions

S.C., R.L., P.D., and N.M. designed the research ; S.C. and R.L. performed the research ; S.C. and R.L. contributed analytic tools ; S.C., R.L., P.D., and N.M. analyzed the data ; and S.C., P.D., and N.M. wrote the manuscript.

6.9 Acknowledgements

This work was funded in part by the Canada Research Chairs program, the Fonds québécois de recherche sur la nature et les technologies (FQRNT), the Natural Sciences and Engineering Reasearch Council of Canada (NSERC) and the Fonds de recherche en santé du Québec (FRSQ). Calculations were done on the computers of Calcul Québec. Financial support of CNRS and the Institut Universitaire de France is also acknowledged.

CHAPITRE 7

SYNTHÈSE DES EXPÉRIENCES ET SIMULATIONS SUR LES DIMÈRES ET PETITS OLIGOMÈRES Aβ

Simulations of Aβ40/42 dimers and higher order assemblies in aqueous solution

Sébastien Côté and Normand Mousseau

Département de Physique and Groupe d'étude des protéines membranaires (GÉPROM), Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec), Canada

Manuscrit ayant servi à l'élaboration des Sections 2.2 et 4 de : Nasica-Labouze, J., P. H. Nguyen, F. Sterpone, O. Berthoumieu, N.-V. Buchete, S. Côté, A. De Simone, A. J. Doig, P. Faller, A. Garcia, A. Laio, M. S. Li, S. Melchionna, N. Mousseau, Y. Mu, A. Paravastu, S. Pasquali, D. J. Rosenman, B. Strodel, B. Tarus, J. H. Viles, T. Zhang, C. Wang, and P. Derreumaux. 2015. Amyloid β protein and Alzheimer's disease : when computer simulations complement experimental studies. *Chem. Rev.* 115 :3518-3563.

7.1 Introduction

Soluble Aβ oligomers [418] as small as the dimer [422] are pathologically relevant to Alzheimer's disease through various cellular pathways [37, 229, 246, 296, 488]. The *in vivo* population of Aβ consists of a variety of alloforms. These originate from various cleavages and mutations of the amyloid precursor protein (APP) and show distinct biophysical properties. A description of the key structural features of the oligomers formed during the distinct aggregation pathways of each alloform is therefore fundamental to identify and design drugs to control or inhibit their formation.

Experiments on various Aβ40/42 alloforms have unveiled a wealth of information on the size distribution and the low-resolution structure of the oligomers. Precise atomistic characterization of the conformational ensembles of Aβ40/42 soluble oligomers remains elusive, however, as oligomers are structurally highly degenerated and aggregation-prone. *In silico* studies are a necessary complement to in vitro experiments for the atomic-level characterization of the structural flexibility of these oligomers.

We first describe the experimental observations on Aβ40/42 oligomers and then relate those to results obtained with various simulation protocols on the dimer and higher order assemblies in aqueous solution. We conclude with a perspective on this challenging system. For comparison purposes, the Aβ sequence is treated as four regions : N-terminal (residues 1-16), central hydrophobic core or CHC (residues 17-21), loop region in the fibrils (residues 22-28) and C-terminal (residues 29-40/42).

7.2 Observations from experiments

A wide range of experiments was designed to probe the structure and early aggregation properties of Aβ40/42 oligomers. Photo-induced cross-linking of unmodified protein (PICUP) followed by gel electrophoresis (SDS-PAGE) has been extensively applied by Teplow and coworkers over many years to unveil the size distribution of the early short-lived metastable Aβ oligomers [53, 54, 402]. A first study showed that Aβ1-40 is mainly composed of monomers to tetramers in rapid equilibrium [54], while Aβ1-42 preferentially forms larger oligomers such as pentamers and hexamers [53]. This difference was linked to the specific roles of Ile-41 and Ala-42 : Ile-41 appears to be essential for facilitating paranucleus formation while Ala-42 enhances the self-association of these paranuclei [53]. Recent PICUP results further suggest that a turn centered at residues Val-36 and Gly-37 of the C-terminal region of Aβ1-42 and absent from Aβ1-40 is responsible for the characteristic features of Aβ1-42 early oligomers [402]. Single-molecule atomic force microscopy (SMFS) tends to confirm this latter observation for the dimer and suggests that the increased dimer stability of Aβ1-42 compared to Aβ1-40 is due to a shift of the primary interaction site from the N-terminal for Aβ1-40 to the C-terminal for Aβ1-42 [519].

Comparing 34 physiologically relevant Aβ sequences, Teplow et al. further characterized the role of various residues [55]. This extensive study suggests, for instance, that Aβ1-40 is mostly sensitive to mutations at Glu-22 and Asp-23 such as E22G (Artic) and D23N (Iowa) in the loop region as well as N-terminal truncations, while Aβ1-42 is most affected by mutations in the CHC region such as A21G (Flemish) as well as C-terminal truncations. Other PICUP results also support the observation that Aβ oligomerization is highly sensitive to perturbations of the N-terminal and C-terminal [303] as demonstrated, for example, by the D7N (Tottori) familial mutation which accelerates the kinetics of transition to β-sheet-rich configurations and promotes

the early formation of higher order oligomers with more α/β structures that are significantly more toxic compared to wild types Aβ1-40 and Aβ1-42 [356].

Secondary structure was probed by circular dichroism (CD) spectroscopy on both cross-linked and uncross-linked populations of early oligomers. Early work suggests that these are mostly disordered with a relatively small amount of β-sheets/β-turns (13-20% for Aβ1-40 and 18-24% for Aβ1-42) and very low α-helix content (3-7% each) [53]. More recent work, by the same group, finds higher secondary structure content for Aβ1-40, with α-helix varying from 9 to 13% and β-sheets going from 24 to 45% between the monomer and the tetramer, with the largest β-sheet gain being between the monomer and the dimer [355]. Interestingly, this latter study also shows that the toxicity increases with the β-sheet propensity and the oligomer order [355].

Other groups confirmed that different Aβ alloforms have distinct early oligomerization. For instance, Chen and Glabe determined, using fluorescence and gel electrophoresis, that Aβ1-40 mostly populates unstable monomeric conformations, while Aβ1-42 samples larger oligomers such as trimers and tetramers that are significantly more stable [93]. Bowers and co-workers developed a new method, ion mobility coupled to mass spectrometry (IM-MS), to investigate the early oligomers of Aβ1-40 and Aβ1-42 [45]. They confirmed that Aβ1-40 dominantly populates monomers, dimers and tetramers, while Aβ1-42 mostly forms dimers, tetramers, hexamers and dodecamers. Moreover, they proposed an assembly mechanism in which the dimer plays a key role and they identified structural differences in the tetramer that rationalize the formation of higher order oligomers by Aβ1-42, but not by Aβ1-40 [45]. The effects of familial mutations such as D7N (Tottori), A21G (Flemish) and E22G (Artic) have also been studied using IM-MS showing that D7N and E22G in Aβ1-40 produce a tetramer conformation that is similar to Aβ1-42's leading to the formation of larger aggregates not populated in wild type (hexamers for D7N and E22G, and dodecamers for E22G), while A21G in Aβ1-40 changes the collapse tetramer of the wild type to a significantly more extended conformation [169]. The equivalent mutations in Aβ1-42 either significantly increase the stability of the dodecamer compared to smaller oligomers (D7N and E22G) or completely change the topology of the hexamer abrogating the formation of the dodecamer (A21G) [169].

Various other experimental techniques have also unveiled quantity of low-resolution structural information on the early oligomers of Aβ [296]. While detailed high-resolution structural

information for a wide range of specific oligomer orders is still lacking, selective labelling nuclear magnetic resonance (NMR) spectroscopy combined with other biophysical methods such as atomic force microscopy (AFM) and Fourier transform infrared (FTIR) spectroscopy has unveiled some structural features of Aβ oligomers : the C-terminal of Aβ1-42 preglobulomers and globulomers is fully protected from solvent exchange and is involved in intra- and intermolecular β-sheets [515], the Aβ1-42 pentamer can form disc-shaped conformations with low secondary structure signals in which the CHC and C-terminal regions are protected from solvent exchange [3], intermolecular interactions and β-sheet organization of Aβ1-42 oligomers that appear off-pathway are different from the fibril [458], and Aβ1-40 populates an heterogeneous ensemble of oligomeric structures co-existing with pre-fibrillar and fibrillar aggregates [390, 451]. Specific intra- and inter-molecular side-chain contacts are however different from one NMR study to the other leading to very different models for the tertiary and quaternary structures of Aβ oligomers, while all agree on them weakly populating in-register parallel β-sheets characteristic of the fibril. New flavours of mass spectrometry (MS) measurements are also being developed to look at the structural features of low molecular weight Aβ oligomers [80, 362, 363, 532].

7.3 Observations from simulations – Dimers

Even for the dimer, computational efforts are such that one must chose between a sufficient sampling with a simplified forcefield or a limited simulation with a finer all-atom potential as discussed in this section. The simulations discussed are summarized in Table 7.I.

7.3.1 Simplified representations

Using the implicit solvent OPEP forcefield, which includes all main chain atoms and a single bead for side chains [311, 314, 498], Derreumaux, Mousseau and co-workers studied the structural ensemble of Aβ1-40, Aβ1-42 and Aβ1-40(D23N) monomers [108] and dimers [109] using HT-REMD simulations [86, 256] and starting from random conformations. These studies find that these three alloforms populate mostly turn/random coil dimers with a β-sheet propensity at the C-terminal region larger than for the monomers. Dimerization is mainly driven by a hydrophobic sequestration through the formation of interchain contacts between CHC/CHC, CHC/C-terminal and C-terminal/C-terminal. As a result, the Aβ1-42 dimer, with its additional residues, is significantly more compact and more structured than Aβ1-40 with larger β-sheet and contact

Table 7.I – Summary of the simulations discussed in sections 7.3 and 7.4.1.

Ref.	Force field	Solvent model	Method	Timescale	Aβ alloforms	Oligomer order
32	OPEP 3.2	Implicit	HT-REMD	1.25 μs × 26 rep	1-40, 1-42, 1-40(D23N)	2
38	CHARMM19	Implicit SASA	T-REMD	0.4 μs × 24 rep × 7 sim	10-40	2
40	CHARMM19	Implicit SASA	T-REMD	0.8 μs × 24 rep × 8 sim	1-40	2
44	PROFASI	Implicit	MC	2×10^{10} steps × 40 sim	1-42 WT, F20E, E22G, E22G/I31E	2
48†	DMD4B–HYDRA	Implicit	DMD	1×10^{7} steps × 8 sim	1-40, 1-42	1–32
49†	DMD4B–HYDRA	Implicit	DMD	6×10^{7} steps × 8 sim	1-40, 1-40(E22G), 1-42, 1-42(E22G)	1–32
53‡	OPLS-AA\L	SPCE	MD	0.05 μs × 1000 sim	1-40, 1-42	2
	OPLS-AA\L	TIP3P	MD	0.05 μs × 1000 sim	1-40, 1-42	2
55	OPLS-AA\L	SPC	T-REMD	0.2 μs × 64 rep	1-42	2
57	AMBER99sb	TIP4P-Ew	T-REMD	0.05 μs × 52 rep	1-42, 1-43	2
58	AMBER99	TIP3P	MD	0.1 μs	1-42	2
59	OPLS-AA\L	TIP3P	MD	0.8 μs	1-40, 1-40(D7N), 1-42, 1-42(D7N)	2
60	GROMOS53A5	SPC	MD	0.05 μs × 10 sim	1-42	2
63	GROMOS96	SPC	MD	0.01 μs	1-40, 1-40(A21G), 1-42, 1-42(A21G)	2
64	OPEP 3.2	Implicit	T-REMD	1.2 μs × 22 rep	17-42	3
48†	DMD4B–HYDRA	Implicit	DMD	1×10^{7} steps × 8 sim	1-40, 1-42	1–32
49†	DMD4B–HYDRA	Implicit	DMD	6×10^{7} steps × 8 sim	1-40, 1-40(E22G), 1-42, 1-42(E22G)	1–32
66†	DMD4B–HYDRA	Implicit	DMD	4×10^{7} steps × 8 sim	3-40, 11-40, 3-42, 11-42	1–32
67	CHARMM19	Implicit SASA	T-REMD	0.8 μs × 24 rep × 8 sim	10-40	4
68	CHARMM19	Implicit SASA	T-REMD	0.8 μs × 24 rep × 8 sim	10-40	4
69†	OPLS-AA\L	Implicit GB/SA	T-REMD	0.2 μs × 5 sim	1-42	1–20

Large scale aggregation simulations resulting in an ensemble of oligomer order sampled and analysed. ‡ The all-atom, explicit solvent simulations in ref. [25] are started from the conformations obtained from the extensive coarse-grained sampling in ref. [473].

propensities at the CHC and C-terminal. The D23N (Iowa) mutation, for its part, causes non-local perturbations of the wild type conformational ensemble by, for instance, increasing the β-sheet propensity at the C-terminal region resulting in new motifs there such as an intermolecular anti-parallel β-sheet that is similar to its fibril topology.

United-atom implicit solvent simulations with the CHARMM19+SASA forcefield [68, 155] using REMD [449] were performed by Klimov and co-workers on the Aβ1-40 and Aβ10-40 monomers and dimers [453, 455, 456]. The authors find that dimerization increases the propensity of extended structure at the expense of helical structure [453]. They also show that main chain hydrogen bonds are not crucial to dimerization as their absence do not significantly change the network of interpeptide contacts [456]. These latter results are presumed to hold for the full-length alloform as they observe that the truncation of the first 9 residues leads to minor changes in the structure of the Aβ1-40 dimer [455]. Moreover, by reshuffling the amino acid sequence of one of the Aβ peptide, they find that it does not impact the molten globule nature of the dimer and its assembly suggesting that it behaves in large part like a molten globule polymer [288].

The effective all-atom implicit solvent model developed by Irbäck and co-workers and coupled to Monte Carlo simulations [197] was used to study wild-type Aβ1-42 monomer [324] and dimer [325] and the effect of three mutations : F20E, E22G (Artic) and E22G/I31E. Here, the dimerization is characterized by an increase of β-sheet content for these four alloforms of Aβ1-

42. The resulting dimers are mostly composed of intramolecular anti-parallel β-sheets, while they marginally populate intermolecular β-sheets, in-register parallel β-sheets and α-helices. The dimer is mostly stabilized by interchain contacts involving residues Phe-19 and Phe-20 of the CHC region in the wild type, E22G and E22G/I31E alloforms, while the F20E mutation destabilizes these contacts reducing the aggregation propensity. The effect of these mutations, which have very different aggregation properties in vitro [55], is rather focused on two turns located in the loop region suggesting that reorganization of this part of the peptide is a potentially critical step in Aβ aggregation.

Finally, structural results from large scale DMD simulations using a 4-bead coarse-grained model developed by Urbanc and co-workers [50, 428, 474] on a system of 32 peptides were used as a starting point for the statistical analysis of dimer conformations [473, 476]. These dimeric conformations are collapsed and disordered with a small content of β-sheets linked by loops and turns, as planar β-sheets are thermodynamically unstable for both sequences [475]. The monomer-to-dimer transition appears driven by an hydrophobic collapse and it is characterized by an increase of β-sheet propensity with the resulting Aβ1-42 dimer having a larger propensity of β-sheets at the CHC and C-terminal compared to Aβ1-40 [473]. These two regions also form the dimer core with large propensity of intermolecular contacts, dominant at the CHC for Aβ1-40 and at the C-terminal for Aβ1-42. All-atom explicit solvent OPLS-AA\L [220, 220] MD stability simulations on Aβ1-40 and Aβ1-42 monomers and dimers starting from the DMD-obtained coarse-grained structures confirm previous DMD results [473] and enable a precise analysis of salt-bridges propensity and free energy landscape [25]. For instance, the formation/breaking of salt-bridges is in dynamical equilibrium as the charged residues are highly accessible to the solvent with D23-K28 and E22-K28 being the most populated salt-bridges with a propensity significantly smaller in the Aβ1-42 dimer than Aβ1-40. The free-energy landscape of the Aβ1-42 dimer is also found to be more complex than Aβ1-40 showing strong differences between these two alloforms already at the dimer.

7.3.2 All-atom explicit solvent simulations

All-atom explicit solvent REMD [449] simulations with the OPLS-AA\L forcefield [220, 220] – which was found to provide a reasonable agreement with NMR results [420] – were recently performed on the Aβ1-42 dimer by Mu and co-workers [529]. They show that the Aβ1-42 dimer

is rather unstructured as it is mostly composed of coils, bends and turns with a small β-sheet propensity that is mainly localized at the C-terminal region. The main contribution to the free energy of dimerization comes from the formation of interpeptide hydrophobic contacts between C-terminal/C-terminal, CHC/CHC and CHC/C-terminal, while electrostatic interactions such as salt-bridges are not very populated in the dimer. In contrast, short 50-ns per-replica all-atom explicit solvent REMD simulations with the AMBER99sb forcefield [190] in the Aβ1-42 and Aβ1-43 dimers suggest that the Aβ1-42 dimer has a flexible C-terminal and, following ab-initio energy calculations, is mainly stabilized by electrostatic interactions [513]. A solvation free energy analysis based on the integral-equation theory of liquid applied to trajectories of relatively short MD dimerization simulations (100 ns) suggests that dimerization occurs through a two-step nucleation-accommodation mechanism : decrease of each monomer solvation free-energy upon their approach followed by structural reorganizations in the newly formed dimer leading to a decrease in the protein internal energy [101].

Due to the sampling cost of unbiased simulations, most other all-atom explicit solvent MD simulation focus rather on the stability of either experimentally derived or theoretically constructed dimeric structures. For instance, OPLS-AA\L [220, 220] MD stability study of the Aβ1-40 and Aβ1-42 dimers starting from the sheet-loop-sheet fibril state suggest that the D7N (Tottori) mutation enhances the aggregation rate by various pathways, as proposed experimentally [169, 356], such as by decreasing the turn propensity at the N-terminal around residues 8-9 and by perturbing salt-bridges half-lives and by increasing the population of fibrillar-like states for Aβ1-40 [482]. Other MD simulations indicate that the association free energy of the Aβ1-42 dimer would be mainly due to the formation of non-polar contacts as the CHC and C-terminal regions are crucial to stabilize the dimer [538], in agreement with results from sampling enhancing techniques [109, 325, 473]. The A21G (Flemish) mutation on the Aβ1-40 and Aβ1-42 dimers starting from the fibril structure of the Aβ1-40 fibril [375] was studied, here with the GROMOS96 forcefield [195, 440]. Their results suggest that this mutation reduces the aggregation rate by destabilizing β-sheets in Aβ1-40, by increasing the flexibility of the CHC region in Aβ1-42 and more significantly in Aβ1-40, and by perturbing the network of salt-bridges such as D23-K28 in both sequences [195].

7.3.3 Towards a converging atomistic model for the dimer ?

Common trends can be extracted from these various simulation protocols applied to study the Aβ40/42 dimers : (1) the monomer-to-dimer conversion of wild type Aβ1-40 [109, 455, 473] and Aβ1-42 [109, 325, 455] is characterized by an increase in extended/β-sheet structures as measured experimentally [355], (2) dimerization of Aβ is mainly driven by an hydrophobic collapse through the formation of intermolecular contacts involving the CHC and the C-terminal regions [25, 109, 325, 473, 529, 538] agreeing with the importance of these regions during aggregation as observed experimentally [3, 55, 515], (3) the Aβ1-42 dimer has larger β-sheet propensity than Aβ1-40 particularly at the CHC and C-terminal regions [25, 109, 473], (4) marginal α-helix propensity for both the Aβ1-40 and Aβ1-42 dimers [25, 109, 325, 473, 529] ; except results from one study [453], (5) the free-energy landscape of the Aβ1-42 dimer is more complex than Aβ1-40 [25, 109], (6) salt-bridges are highly accessible to the solvent with D23-K28 and E22-K28 being the most populated ones and are more present in the Aβ1-40 dimer than Aβ1-42 [25, 109], and (7) point mutations in the CHC region such as F20E, in the loop region such as D23N (Iowa) or E22G (Artic) and in the N-terminal such as D7N (Tottori) cause important local and non-local conformational changes [109, 325, 455, 482]. These trends suggest that Aβ alloforms have important structural differences already at the dimer that could account for their very different oligomerization pathways and toxicity potency as observed experimentally [45, 53, 55, 169, 303, 356, 402, 519]. Detailed comparison between these simulation results, based on diverging protocols, however, identifies quantitative and qualitative differences concerning the net and the individual propensities of secondary (see Figure 7.1 and Table 7.II) and tertiary/quaternary structures. More precise experimental structural measurements on the monomer and higher order oligomers as well as standardized simulations with various description levels will help converge on the most relevant structures.

7.4 Observations from simulations – Higher order assemblies

Simulations of higher order assemblies are even more challenging as the complexity of the conformational space scales rapidly with the number of peptides. Coarse-grained and all-atom models coupled to implicit solvent schemes can be used to access the relevant timescales for their folding by reducing the complexity of the molecular system. All-atom explicit solvent MD simulations, for their part, are limited to the study of the stability of preformed topologies. Be-

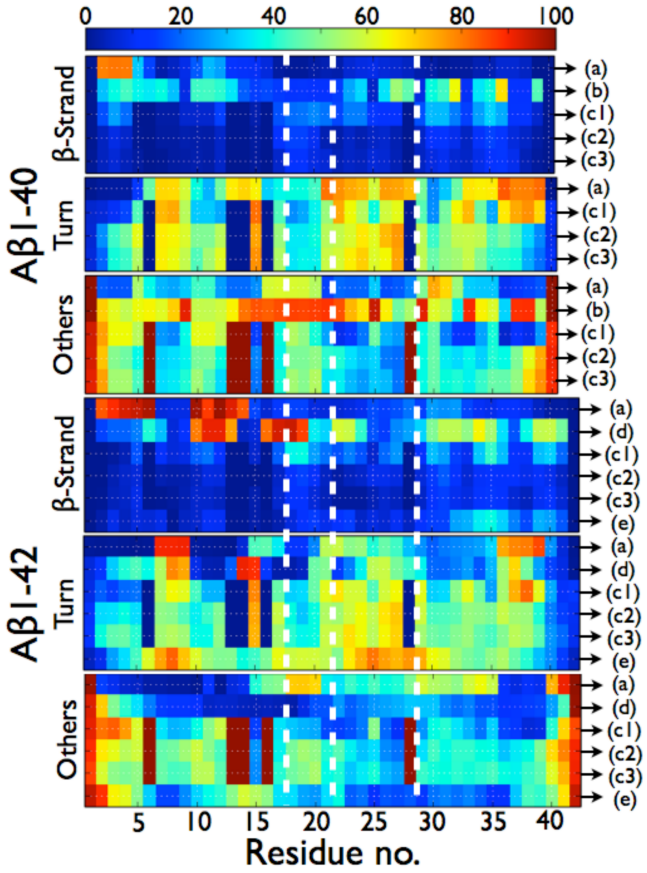


Figure 7.1 – Per residue probabilities of β-strand, turn and others for the Aβ1-40 and Aβ1-42 dimers as computed by different simulation protocols : (a) coarse-grained implicit solvent OPEP HT-REMD [109], (b) united atom implicit solvent CHARMM19+SASA REMD [455], (c1) coarse-grained implicit solvent DMD [25], (c2) all-atom explicit solvent OPLS-AA/SPCE MD [25], (c3) all-atom explicit solvent OPLS-AA/TIP3P MD [25], (d) effective all-atom implicit solvent MC [325], and (e) all-atom explicit solvent OPLS-AA REMD [529]. All secondary structure probabilities shown were computed using STRIDE ; except for (b) for which the secondary structures are determined using information on the ϕ and ψ dihedral angles only, without consideration of the h-bond network. As such the β-sheet probabilities for (b) must be considered as extended conformations and no turn probability can be determined. The vertical dotted white lines delimit the four regions : N-terminal (residues 1-16), CHC (residues 17-21), loop region in the fibril (residues 22-28), and C-terminal (residues 29-40/42).

low, we focus our attention on non-fibrillar oligomeric aggregates. The aggregation simulations discussed are summarized in Table 7.I.

7.4.1 Aggregation

The coarse-grained implicit solvent OPEP forcefield developed by Derreumaux and co-workers for amyloid and non-amyloid proteins [311, 498] was used to study the structural ensemble of

Table 7.II – Overall secondary structure probabilities for the Aβ1-40 and Aβ1-42 dimers.

Ref.	Aβ1-40				Aβ1-42			
	Helix	β-Strand	Turn	Coil	Helix	β-Strand	Turn	Coil
	%	%	%	%	%	%	%	%
15†	10.5	38.6	–	50.9	–	–	–	–
32	1.3 ± 0.1	12.6 ± 0.1	50.7 ± 0.1	35.4 ± 0.1	4.4 ± 0.1	30.8 ± 0.1	32.4 ± 0.1	32.4 ± 0.1
44	–	–	–	–	0.9 ± 0.1	42.9 ± 0.7	33.0 ± 0.4	23.2 ± 0.8
49	0.1 ± 0.1	13.6 ± 1.6	40.6 ± 4.1	37.5 ± 5.0	0.0 ± 0.0	15.7 ± 1.9	39.2 ± 3.7	39.0 ± 4.9
53‡	0.5 ± 0.1	5.5 ± 0.8	43.5 ± 3.6	46.8 ± 4.1	0.9 ± 0.2	6.6 ± 0.8	40.1 ± 3.2	48.0 ± 3.8
	0.9 ± 0.1	4.8 ± 0.6	42.5 ± 3.6	47.9 ± 4.0	0.8 ± 0.1	5.6 ± 0.7	41.6 ± 3.1	47.7 ± 3.6
55	–	–	–	–	8.4 ± 0.7	11.9 ± 0.6	51.2 ± 1.0	28.5 ± 0.8

Values from simulations are computed using STRIDE. † Experimental measurements using CD spectroscopy on extracted PICUP cross-linked dimers. ‡ Values for the simulations using OPLS-AA\L with SPCE and TIP3P are respectively shown first and second.

the Aβ17-42 trimer [87] using REMD simulations [86]. The Aβ17-42 trimer adopts globular conformations mostly populating turns and random coils with low propensity of β-sheets. As for the dimer, the trimer is mainly stabilized by intermolecular contacts between the CHC and C-terminal regions.

The 4-bead coarse-grained model developed by Urbanc and co-workers for DMD simulations [474] has been used to extensively study the aggregation of various alloforms of Aβ40/42 [50, 316, 428, 473, 476, 517]. Their 32-peptide simulations reproduce qualitatively the main features of the early oligomer size distribution of Aβ1-40 and Aβ1-42 as measured experimentally [53]. Overall, these oligomers, as for the dimer, form rather amorphous aggregates with a low propensity of β-sheet that are stabilized by intermolecular contacts between the CHC and C-terminal regions [473, 476]. The main differences between Aβ1-42 and Aβ1-40 oligomers is the larger β-sheet propensity at the C-terminal region as well as the larger flexibility and solvent-exposure of the N-terminal of the former alloform. Truncation of the N-terminal residues (Aβ3-40, Aβ3-42, Aβ11-40 and Aβ11-42) is found to shift the oligomer size distribution towards larger oligomers [316] as observed experimentally [55]. Moreover, the fact that the N-terminal of the Aβ3-40 and Aβ3-42 variants is more flexible than Aβ1-40/42 and Aβ11-40/42 could be related to the increased toxicity of Aβ3-40/42 and their pyroglutamated isoforms Aβ3(pE)-40/42 compared to the full-length sequence [316].

United-atom implicit solvent simulations with the CHARMM19+SASA forcefield [68, 155] using REMD [449] have been performed by Klimov and co-workers on the Aβ10-40 tetramer [240, 241]. Their results also show a rather amorphous tetramer that is structurally similar

to the dimer : they do not populate fibril-like conformations, and their secondary structures are very similar.

All-atom MD simulations with the OPLS-AA\L forcefield [220, 220] and the implicit solvent scheme GB/SA by Barz et al. on a system of 20 Aβ1-42 revealed that the early aggregation pathways are very diverse and are dominated by unstructured oligomers [24] with an oligomeric size distribution similar to previous experimental results [53]. These conformations are mostly characterized by strong intermolecular interactions involving the C-terminal and CHC regions. A maximum flow transition network (MTN) analysis applied on the observed conformations during oligomerization further unveiled key preferential aggregation pathways such as : the trimer serves as building block for the hexamer, while the dimer preferentially aggregates into tetramer and octomer.

7.4.2 Stability of preformed oligomers

A number of simulations were performed to test the stability of preformed oligomeric assemblies. Starting from the crystal structure of the Aβ-IgNAR chimeric tetramer consisting of a core stabilized by anti-parallel β-sheets between the Aβ part [447], the MD simulations of Scoher et al. using the AMBER99sb forcefield [190] and explicit solvent suggest that this conformation is stable for Aβ17-42, but not for Aβ17-40 due to a stability increase of the β-sheets at the C-terminal region [435]. Using the all-atom CHARMM27 forcefield [6, 299] with explicit solvent, Yu and Zheng performed MD simulations to evaluate the stability of two classes of Aβ17-42 12-mers circular globulomers assembled from monomeric and dimeric subunits having the sheet-loop-sheet topology observed in the fibril [516]. These globulomers sample polymorphic structures with a common structural feature : the C-terminal forms the core of the aggregates while the rest of the sequence is exposed to the solvent. Ma and Nussinov showed, using MD simulations with the CHARMM27 forcefield [299] and explicit solvent, that two disk-like topologies inspired from experimental observations [3, 515] for the Aβ1-42 pentamer are highly flexible resulting in conformations that are very different from the fibril topology [297]. Guy and co-workers performed MD simulations using CHARMM27 forcefield [299] and DMD simulations using a 4-bead coarse-grained model [474] on nine Aβ1-42 hexamer topologies in which the C-terminal residues are involved in a six-stranded parallel or anti-parallel β-barrel topology forming the hydrophobic core of these folds [421, 518]. Their simulations show that

most of their proposed hexameric assemblies are very stable suggesting that they are viable topologies for the polymorphic Aβ hexamer. Nussinov and co-workers also investigated, using solvent explicit MD simulations with the CHARMM27 forcefield [299], the stability in aqueous solution of single- and double-layered annular Aβ topologies with an inner diameter of 2 nm built from monomeric subunits having the sheet-loop-sheet motif [534]. They observe that the single-layered annular oligomer is intrinsically unstable, while the double-layered annular oligomer for which the C-terminal region of each layer is facing each other is stable.

Overall, the main driving force for the formation of larger oligomers and their stability remains, as for the dimer, the sequestering of hydrophobic residues of the CHC and C-terminal regions away from the solvent [24, 50, 87, 316, 421, 428, 516–518]. Simulations starting from preformed topologies suggest that small oligomers such as the Aβ42 hexamer [421, 518] and tetramer [435] can populate stable structured folds, as recently suggested from experimental results [45, 447], the formation of which likely involves large conformational changes requiring to pass through many metastable states. Simulations starting from a disordered state, on the other hand, produce mostly disordered Aβ oligomers having few secondary structure elements, also in agreement with experiments [53, 355], and sharing strong structural similarities with the dimer [24, 50, 87, 240, 241, 316, 428, 516, 517]. This suggests that sampling, starting form a disordered state, is still insufficient to converge towards rarer more ordered structures.

7.5 Perspective

Simulations have an important role to play in the identification of the out-of-equilibrium oligomeric structures. Results published over the last few years show that it is now just getting possible to study numerically the aggregation process. Key advancements in structural characterization of amyloid oligomers also provide new grounds for quantitative comparison [345]. While most simulations agree on a few fundamental points, they still wildly differ on a number of aspects due, mostly, to the difficulties associated with correctly sampling these large and slow systems. More comparisons between simulations are clearly required as well as multiscale approaches such as that followed by Urbanc et al. [25], that couple cheap potential for fast sampling and more reliable forcefields for optimizing the coarse-grained models. Carefully prepared thermodynamical simulations, such as REMD and Metadynamics, could also provide important information regarding the free-energy landscape of low oligomer weights such as the

dimer. Clearly, new coarse-grained potentials that could cover the gap between the current CG and complete atomic description are still needed.

CHAPITRE 8

PRÉAMBULE AUX ÉTUDES SUR LA PROTÉINE HUNTINGTINE

Huntingtine est une protéine de plus de 3000 acides aminés qui possède plusieurs rôles physiologiques dans la cellule [133, 424]. Elle est, entre autres, essentielle à l'embryogenèse [143] et elle est impliquée dans la transcription ainsi que dans le transport intracellulaire d'organites et de vésicules [10, 79, 298, 535]. Or, il s'agit surtout de son association avec la maladie de Huntington qui en fait une protéine très étudiée [26]. La maladie de Huntington est une maladie neurodégénérative caractérisée par d'importants troubles moteurs et cognitifs qui sont causés par une atrophie du cerveau, particulièrement des ganglions de la base. Il s'agit d'une maladie génétique à transmission autosomique dominante dont la progression des symptômes peut être ralentie par certains médicaments, mais dont il n'est pas possible de guérir à ce jour. L'incidence de cette maladie est moins prononcée que pour la maladie d'Alzheimer, environ 8 cas sur 100 000 personnes, car elle est transmise uniquement de manière génétique.

L'origine génétique de la maladie est la présence d'un segment de codons CAG consécutifs dans le fragment d'ADN codant pour la protéine huntingtine qui se traduit en un segment de glutamines consécutives lors de sa transcription [168, 358, 539]. Ce segment, appelé Q_N, est situé dans la région N-terminal de la protéine (Figure 8.1). Il contient un nombre anormal de répétitions qui dépasse 35 glutamines (35 codons CAG) lorsqu'une personne est atteinte par la maladie de Huntington. De plus, l'agressivité de la maladie augmente en fonction de la longueur de ce segment. Cette maladie est caractérisée par une mauvaise localisation de huntingtine dans la cellule ainsi que par son mauvais repliement et son agrégation en fibres amyloïdes. Les chemins cellulaires ainsi perturbés sont nombreux [38, 395, 540].

Malgré le fait que huntingtine soit une très grosse protéine, le premier exon de sa séquence est suffisant pour produire le phénotype de la maladie de Huntington *in vivo* [120, 304] et *in vitro* [99, 350, 526]. Il peut être créé *in vivo* par la protéolyse de huntingtine et il est composée de seulement un peu plus de 100 acides aminés sur les plus de 3000 que compte la séquence entière. Le premier exon est séparé en quatre segments en partant du N-terminal : le segment amphiphile Htt17 ayant 17 acides aminés suivi par le segment amyloïdogénique Q_N ayant un nombre variable de glutamines, le segment P_{11} composé de onze prolines et un segment riche en prolines.

Le segment Q_N est responsable de la formation des fibres amyloïdes dont la structure et les processus d'agrégation ont été étudiés en détail par diverses méthodes expérimentales [414, 501]. Il est reconnu, entre autres, que les segments adjacents à Q_N, en particulier Htt17, influencent grandement son agrégation [275, 405, 459], tandis que seul Q_N participe aux feuillets-β formant le coeur de la fibre amyloïde [187].

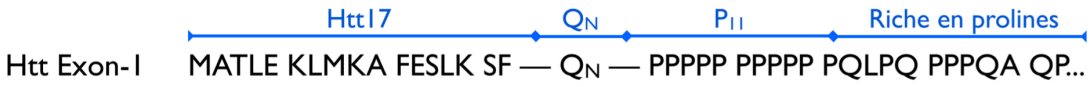


Figure 8.1 – Illustration de la séquence d'acides aminés du N-terminal de la protéine huntingtine. La nomenclature des segments – Htt17, Q_N et P_{11} – utilisée lors des discussions sur huntingtine est aussi présentée. Le segment Htt17 est crucial aux interactions membranaires de huntingtine. Q_N est un segment de glutamines consécutifs responsable de la forme de fibres amyloïdes.

Le segment Htt17 est aussi fondamental aux fonctions physiologiques de huntingtine. Grâce à son caractère amphiphile, il sert d'ancre membranaire contrôlant la localisation de huntingtine dans la cellule entre le noyau, le cytosol, les mitochondries et l'appareil de Golgi [16, 302, 398, 536, 537]. Il peut subir des modifications post-traductionnelles qui changent la localisation et la fonction de huntingtine [4, 14, 175, 443, 463], et il peut interagir directement avec les chaperonnes TRiC éliminant le mauvais repliement et l'agrégation de huntingtine [457].

Le N-terminal de la protéine huntingtine adopte des structures principalement désordonnées qui fluctuent de façon importante en solution [16, 459, 504]). Il est donc très difficile d'observer leurs structures avec des méthodes expérimentales à haute résolution. En effet, la RMN ne génère pas un nombre suffisant de contraintes pour développer un modèle tri-dimensionnel de la protéine [459]. Elle a cependant permis d'identifier le fait que les premiers acides aminés de la séquence auraient une propension plus grande à former une structure hélicoïdale. La plupart des informations structurelles proviennent donc du dichroïsme circulaire (CD) montrant que son ensemble structurel contient une population d'hélices-α [16, 209, 459, 504]. En environnement membranaire, de fortes indications suggèrent que Htt17 adopte une hélice-α, mais aucune structure tridimensionnelle ne confirmait cela [16] alors que nous débutions nos simulations. De plus, la détermination de l'ensemble structurel du N-terminal de huntingtine est complexifié par sa propension à s'assembler selon des chemins d'agrégation variés (Figure 8.2).

C'est à ce niveau que les méthodologies *in silico* que nous avons appliquées complètent les

résultats expérimentaux en palliant ce manque d'information sur la structures du N-terminal de la protéine huntingtine, pourtant au coeur de ses fonctions physiologiques et pathologiques. Les chapitres suivants ont été ordonnés selon la taille des systèmes (Figure 8.2) : le repliement des monomères Htt17, Htt17Q_{17} et Htt17$Q_{17}P_{11}$ est présenté au Chapitre 9, la stabilité et la croissance de nanotubules de Htt17Q_{30} et Htt17Q_{40} sont présentées au Chapitre 10, et les interactions membranaires de Htt17 et Htt17Q_{20} sont présentées aux Chapitres 11 et 12. Nous élaborons ci-dessous la rationnelle derrière chaque étude, en ordre chronologique de leur publication.

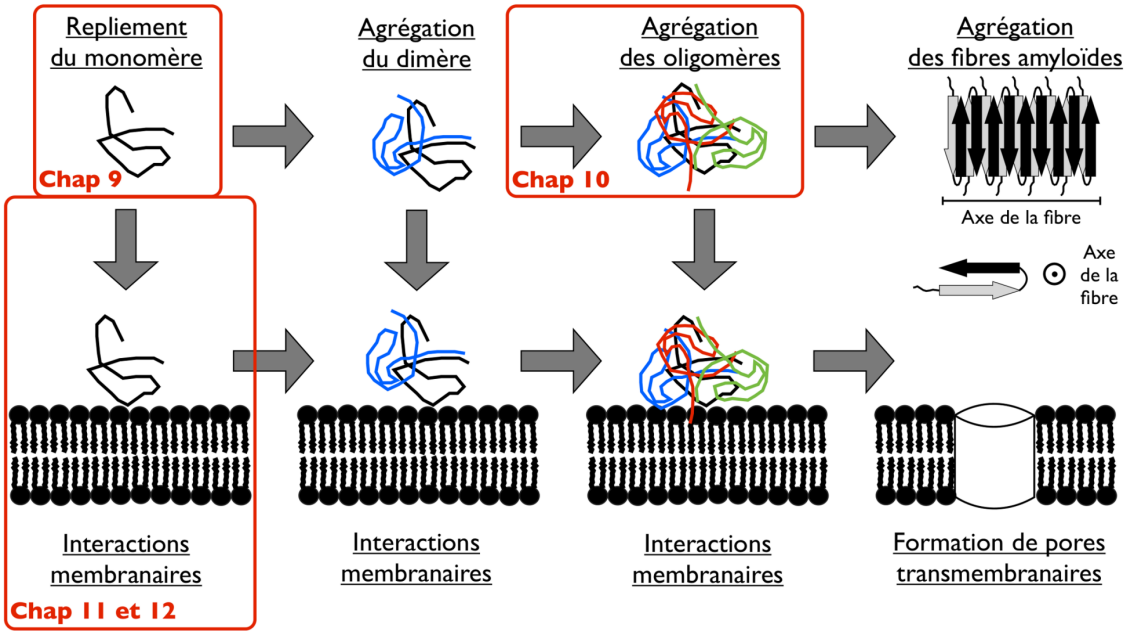


Figure 8.2 – Illustration des chemins d'agrégation du N-terminal de la protéine huntingtine. Les systèmes étudiés aux Chapitres 9, 10, 11 et 12 sont indiqués.

Des simulations avec le modèle gros-grain OPEP [311] et la méthode d'échanges de répliques (REMD) [181, 449] suggèrent que le segment amyloïdogénique Q_N forme des structures nanotubulaires stables lorsque le nombre de glutamines dépasse 36 [255]. Puisqu'il y a peu de données sur la structure de ce segment lors des premières étapes d'agrégation tant du côté expérimental que du côté des simulations, nous avons décidé d'évaluer plus en profondeur la pertinence de ce chemin d'agrégation potentiel. Pour ce faire, nous avons évalué la stabilité et la croissance de ces structures en présence explicite des molécules d'eau du solvant (TIP4P) [216] avec le modèle tout-atome OPLS-AA\L [220] (Chapitre 10). La paramétrisation et la validation de ce modèle sont présentées à la Section 3.1.2. La dynamique moléculaire est utilisée, car le

coût computationnel de la méthode REMD n'est pas raisonnable pour ces systèmes étant donné la taille des structures nanotubulaires simulées en présence d'une représentation explicite du solvant.

Ensuite, nous avons caractérisé les interactions responsables de l'interaction de Htt17 avec une membrane de phospholipides (Chapitres 11 et 12). L'expérience montre que Htt17 se localise sur des membranes lipidiques et certaines indications suggèrent qu'elle adopte la structure d'une hélice-α [16, 230, 232], mais aucun modèle structurel de Htt17 sur une membrane n'était publié lorsque nous avons débuté ce projet. Dans ces simulations, les interactions protéine–protéine sont modélisées avec le champ de force AMBER99sb*-ILDN [382] et le modèle d'eau TIP3P [215], car nous avons observé que cette combinaison semble être la plus appropriée pour décrire la structure de Htt17 en solution aqueuse selon une comparaison avec les mesures expérimentales de dichroïsme circulaire (DC) [16, 459, 504]. De plus, tel que présenté à la Section 3.1.1, ce modèle est suffisamment précis pour décrire adéquatement la stabilité et le repliement d'un ensemble varié de protéines [28, 98, 279, 382, 383] dont des petits peptides désordonnés adoptant transitoirement des hélices-α [46, 380]. Les interactions protéine–lipide et lipide–lipide sont quant à elles décrites par le modèle SLIPIDS [203–206, 361] ou par le modèle BERGER–AMBER [102]. Ces derniers décrivent adéquatement les propriétés structurelles des membranes dans l'ensemble NPT,[1] l'énergie libre du transfert d'acides aminés de l'eau au cyclohexane ainsi que les propriétés dynamiques de plusieurs systèmes protéine–membrane.

Nous avons, dans un premier temps, analysé les étapes d'insertion de Htt17 et Htt17Q_{20} sur une membrane de phospholipides en utilisant la dynamique moléculaire (Chapitre 11). Nous avons aussi confirmé la stabilité de structures hélicoïdales déjà insérées dans la membrane avec une orientation parallèle au plan de la membrane. Dans un deuxième temps, à la suite d'une prise de contact avec un groupe d'expérimentateurs travaillant aussi sur les interactions Htt17–membrane, nous avons caractérisé plus exhaustivement la structure, l'orientation et les interactions qu'adopte Htt17 sur une membrane de phospholipides (Chapitre 12). Nos collaborations ont permis de déterminer la structure du Htt17 en présence de micelles avec la RMN en solution [319], son orientation sur une membrane composée de POPC avec la RMN en état solide [318] ainsi que sa profondeur d'insertion sur plusieurs membranes de phospholipides avec le *quenching* de la fluorescence [317]. Le but de notre nouvel ensemble de simulations est de faire

[1] Au moment de débuter ces simulations, il n'y avait pas de modèle AMBER pour les membranes de phospholipides décrivant adéquatement leurs propriétés dans l'ensemble NPT. Ce modèle existe maintenant [132].

le pont entre ces résultats expérimentaux obtenus dans divers environnements avec différentes techniques. Afin de s'assurer de la validité et de la convergence de nos résultats, nous avons effectué plusieurs dynamiques moléculaires (DM) et échanges de répliques Hamiltonien avec le protocole REST2 (HREX) [492] – décrit à la Section 3.2.3 – en partant de structures insérées dans la membrane. Deux structures initiales pour Htt17 sont utilisées dans nos simulations : elle est complètement hélicoïdale ou alors elle correspond à la structure RMN de nos collaborateurs. De plus, nous avons simulé le repliement en hélice-α du Htt17 sur une membrane en partant de la structure RMN obtenue en présence de micelles. Nos résultats combinés avec les résultats de nos collaborateurs ont permis l'élaboration d'un modèle plus complet sur la structure, l'orientation et les interactions de Htt17 et Htt17Q_N sur une membrane de phospholipides.

Finalement, nous avons caractérisé l'ensemble structurel du monomère de Htt17, Htt17Q_{17} et Htt17$Q_{17}P_{11}$ en solution aqueuse. Le but de cette étude est de mieux caractériser les motifs qui pourraient promouvoir leur agrégation par rapport aux modèles expérimentaux proposés [114, 210, 457] ainsi que leurs interactions membranaires par rapport aux résultats expérimentaux [72, 319] et de nos simulations [106, 111] sur leur structure. Il existe en effet peu de simulations sur ces fragments avec des modèles tout-atome en représentant explicitement les molécules d'eau. Dans nos simulations, les interactions sont décrites par les modèles AMBER99sb*-ILDN pour la protéine et TIP3P pour les molécules d'eau pour les mêmes raisons qu'énoncées précédemment. Afin d'échantillonner exhaustivement l'ensemble structurel accessible à ces fragments, nous avons décidé d'utiliser la méthode métadynamique [20] couplée à la méthode d'échanges de répliques Hamiltonien avec le protocole REST2 (HREX) [492] donnant la méthode hybride HREXMetaD [74]. Ces méthodes sont respectivement présentées aux Sections 3.2.4 et 3.2.3. La MétaD utilise un nombre restreint d'une à trois variables collectives pour favoriser l'exploration du paysage énergétique accessible à la protéine. Dans le cas du repliement d'une protéine, ceci est loin d'être suffisant, ce qui biaise du même coup la reconstruction du paysage énergétique par la méthode. Dans ce cas, il est nécessaire de combiner la MétaD avec une autre méthode d'échantillonnage avancée comme HREX. Nos simulations sont finement comparées avec les résultats expérimentaux de DC et RMN afin de confirmer la validité de l'ensemble structurel observé pour Htt17. Une fois cela fait, nous quantifions l'impact de l'ajout des régions Q_N et P_{11} sur cet ensemble structurel.

CHAPITRE 9

MONOMÈRE DU N-TERMINAL DE HUNTINGTINE

Free-energy landscape of the amino-terminal fragment of huntingtin in aqueous solution

Vincent Binette[△], Sébastien Côté[△], and Normand Mousseau

Département de Physique and Groupe d'étude des protéines membranaires (GÉPROM), Université de Montréal, Montréal (Québec), Canada

△ Vincent Binette and Sébastien Côté contributed equally to this work.

Manuscrit soumis au Biophys. J. : Binette, V., S. Côté, and N. Mousseau. 2015. Free-energy landscape of the amino-terminal fragment of huntingtin in aqueous solution.

9.1 Abstract

The first exon of Huntingtin – a protein with multiple biological functions whose misfolding is related to Huntington's disease – modulates its localization, aggregation and function within the cell. It is composed of a 17-amino-acid amphipathic segment (Htt17), an amyloidogenic segment of consecutive glutamines (Q_N), and a proline-rich segment. Htt17 is of fundamental importance : it serves as a membrane anchor to control the localization of huntingtin, it modulates huntingtin's function through posttranslational modifications, and it controls the self-assembly of the amyloidogenic Q_N segment into oligomers and fibrils. The conformational ensemble of the Htt17 monomer as well as the impact of the polyglutamine and proline-rich segments remain, however, mostly uncharacterized at the atomic level due to their intrinsic flexibility. Here, we unveil the free energy landscape of Htt17, Htt17Q_{17} and Htt17$Q_{17}P_{11}$ using Hamiltonian replica exchange combined to well-tempered metadynamics. We characterize the free energy landscape of these three fragments in terms of a few selected collective variables. Our extensive simulations reveal that the free energy of Htt17 is dominated by a broad ensemble of configurations that concurs with solution NMR chemical shifts. Addition of Q_{17} at its carboxy-terminus reduces the extent of the main basin to more extended configurations of Htt17 with lower helix propensity. Also, the aliphatic carbons of Q_{17} partially sequester the non-polar amino acids of

Htt17. For its part, addition of $Q_{17}P_{11}$ shifts the overall landscape to a more extended and helical Htt17 stabilized by interactions with Q_{17} and P_{11}, which are almost exclusively forming a PPII-helix, as well as by intramolecular H-bonds and salt-bridges. Our characterization of Huntingtin's amino-terminus provides insights on the structural origin of its ability to oligomerize and interact with phospholipid bilayers, processes closely linked to the biological functions of this protein.

9.2 Introduction

Huntingtin is a large ubiquitous protein of more than three thousands amino acids [133, 424]. It is essential to embryonic development [143], it interacts with many proteins through its 36 HEAT repeats [10, 298], it is involved in intracellular organelles and vesicular trafficking [79] as well as transcription and axonal transport [535]. The exon 1 of huntingtin – consisting of an amphipathic sequence of 17 amino acids (Htt17), an amyloidogenic polyglutamine region (Q_N), and a segment of 36 amino acids rich in prolines – is closely linked to Huntingtin's functions. This segment contains a nuclear export sequence that controls the localization of huntingtin between the cytoplasm and the nucleus [302, 536, 537], it can undergo posttranslational modifications affecting the localization and function of huntingtin [4, 14, 175, 443, 463], and it is responsible for the localization of huntingtin to the mitochondria and the Golgi [16, 398].

Over the past years, Huntingtin attracted considerable attention as it is an amyloid protein associated to Huntington's disease, an autosomal dominant genetic disorder [26]. Its assembly into amyloid fibrils is triggered *in vivo* by the expansion of the consecutive segment of glutamines at its first exon above a specific threshold. This characteristic behaviour, which is shared by at least 10 other proteins, is termed the polyglutamine/CAG repeat disorder and is associated to several disorders [168, 358, 539]. More specifically to Huntington's disease, the huntingtin protein misfolds, self-assembles, and mislocalizes in the cell when the Q_N region has more than 36 repeats causing deleterious effects by gain- and loss-of-function through various nuclear and extranuclear pathways [38, 395, 540]. Huntingtin amino-terminus fragments, which can be generated *in vivo* by proteolytic cleavage, are found in post-mortem brain tissue [134] and are involved in the pathogenesis of Huntingtin's disease [242, 392]. The first exon, more precisely, is closely linked to the cytotoxicity as it is sufficient to cause Huntington's phenotype both *in vivo* [120, 304] and *in vitro* [99, 350, 526]. This segment also controls the toxicity, localization and clearance of

mutant huntingtin through posttranslational modifications [4, 14, 175, 443, 463]. It furthermore interacts with the TRiC chaperonin mainly through its Htt17 segment suppressing the misfolding and aggregation of huntingtin [457].

The neighboring regions of Q_N in the first exon are crucial to control its misfolding and amyloidogenesis [501]. In fact, the aggregation of full-length huntingtin exon 1 is very similar to that of Htt17$Q_N P_{10}$ showing the importance of the amino acids right next to Q_N [405]. For instance, the presence of the Htt17 segment accelerates the fibrillation kinetics of Q_N [275, 459], while the P_N segment decelerates it [51, 117]. The nucleus size as well as the overall aggregation pathways of Q_N are also strongly affected by the presence of Htt17 [114, 209, 210, 457]. Some experimental results indicate that it causes the aggregation to split into two main pathways in direct kinetic competition : it proceeds either (i) through the formation of α-helical tetrameric bundles of Htt17 that increase the local concentration of Q_N favoring the nucleation of beta-sheeted structures in it, or (ii) through an unfavorable nucleation in the monomeric Q_N [210]. The Htt17 segment could also facilitate the formation and stability of β-sheeted structures in Q_N by interacting directly with it [110, 457]. Others suggest a less direct role for Htt17 where it could destabilize non-fibrillar aggregates by reducing the entanglement of Q_N [403], thus accelerating the formation of fibrils [114].

Given the importance of Htt17, a characterization of its conformational ensemble at the monomer level could shed light on the atomistic features responsible for its aggregation. Experiments suggest that Htt17 samples transient helical configurations in aqueous solution as circular dichroism data indicates the presence of helical structures [16, 319, 459, 504] and as solution NMR suggests no stable secondary structure motif [459]. Due to its intrinsic flexibility and the absence of stable secondary structure motif, the Htt17 monomer yields too few NMR constraints in aqueous solution to extract any viable three-dimensional structural model [459]. A X-rays model of the chimeric maltose-binding protein – huntingtin exon 1 (MPB-Htt17Q_{17}-Ex1) protein also suggests that Htt17 can adopt helical structures, while the Q_N region is mostly disordered and the P_{11} is a polyproline type-II helix [239]. Few notable simulations on Htt17 complemented these experimental results by describing, to some extend, its conformational ensemble either using (i) all-atom, explicit solvent simulated tempering molecular dynamics [233], (ii) all-atom, implicit solvent Monte Carlo [504], or (iii) all-atom, explicit solvent bias-exchanged metadynamics [401] simulations. All agree that Htt17 samples a broad ensemble of helix/coil structures.

Other simulations were aimed at characterizing the overall effect of increasing the Q_N length in the context of huntingtin exon 1 [137, 260, 504].

The Htt17 is also crucial for the localization of huntingtin in the cell, in part, through direct membrane interactions [16, 398]. More recently, the structure of Htt17 in the presence of DPC micelles has been resolved using solution NMR : it is an alpha-helix from residues 6 to 17, while the rest of the sequence is disordered and highly flexible [318, 319]. Results from solid-state NMR [318] and Hamiltonian replica-exchange all-atom simulations [106] indicate that Htt17 is also an alpha-helix in the context of a membrane bilayer. As the formation of alpha-helical structures in Htt17 prior to its binding seems to favor its membrane partitioning [111], understanding the conformational ensemble of Htt17 in aqueous solution could then unveil motifs beneficial to membrane-binding. In the context of exon 1, the effect of Q_N and P_{11} on the occurrence of such motifs could explain their modulation of Htt17 binding affinity as observed experimentally [72].

Focusing on the identification of Htt17's structural features at the origin of its oligomerization and membrane partitioning, we investigate the free energy landscape of the monomeric Htt17 using all-atom, explicit solvent Hamiltonian replica-exchange metadynamics. Such simulation protocol favors the correct sampling of the entire conformational space physically available to the protein. We also quantify the effect of adding the amyloidogenic Q_N region as well as the P_{11} segment on the global free energy landscape of Htt17 as well as on significant atomistic changes. Overall, such detailed information is necessary to rationalize the importance of Htt17, and it paves the way for the investigation of the oligomerization and membrane binding processes per se using such a similarly stringent simulation protocol.

9.3 Materials and Methods

In this study, we use Hamiltonian replica exchange (HREX) and parallel tempering (PT) combined with well-tempered metadynamics (MetaD) [20, 61, 74, 112, 259] simulations to investigate the free energy landscape of the 17-amino-acid amino-terminus segment (Htt17) of the huntingtin protein in aqueous solution. We also quantify the impact of adding the amyloidogenic polyglutamine (Q_{17}) and the polyproline (P_{11}) segments on Htt17's free energy landscape. The amino acid sequence of Htt17 is MATLEKLMKAFESLKSF and an amidated carboxy-terminus is used for all peptide constructs. All simulations are summarized in Table 9.I. We focus on the HREXMetaD simulations in the main text, while the PTMetaD simulations are presented in the

Supporting Material.

Table 9.I – Summary of the performed simulations.

Simulations	Type	Initial conf.	Time per replica μs	Time μs
Htt17_nmr	HREXMetaD	NMR	0.9×16	14.4
Htt17_coil	HREXMetaD	coil	0.9×16	14.4
$Htt17Q_{17}$	HREXMetaD	NMR/coil	0.9×24	21.6
$Htt17Q_{17}P_{11}$	HREXMetaD	NMR/coil/coil	0.9×24	21.6
Htt17_nmr_remd	PTMetaD	NMR	0.9×64	57.6
Htt17_coil_remd	PTMetaD	coil	0.9×64	57.6

All simulations are done in the NVT ensemble in a rhombic dodecahedron periodic cell ($\alpha = 60^\circ$; $\beta = 90^\circ$; $\gamma = 60^\circ$; $a = b = c = 5.35$ nm and 3500 water molecules for Htt17, $a = b = c = 6.80$ nm and 7000 water molecules for $Htt17Q_{17}$, and $a = b = c = 8.12$ nm 10000 water molecules for $Htt17Q_{17}P_{11}$). We combined well-tempered metadynamics (MetaD) to two other sampling enhancing simulation types : Hamiltonian replica-exchange (HREX) and parallel-tempering (PT). The simulations on Htt17 are started from two different initial configurations : a fully random coil structure and its NMR model in the presence of DPC detergent micelles (PDB 2LD2). The latter configuration is disordered from residues 1 to 5 and an α-helix for the rest of the sequence [318, 319]. In the initial state of the simulations on $Htt17Q_{17}$ and $Htt17Q_{17}P_{11}$, Htt17 is taken as the NMR model, while Q_{17} and P_{11} are completely disordered. We focus on the HREX simulations in the main text, and we present the PT simulations in the Supporting Material.

Simulations protocols. Our molecular dynamics simulations are done with the Gromacs package version 4.6.5 [42, 185, 387, 478] combined with the PLUMED plug-in version 2.0.2 [467] to perform the well-tempered MetaD [20] and HREX [74] parts of our simulations, as described below. We use the all-atom forcefield AMBER99sb*-ILDN [46, 190, 382] as it offers helix/coil-balanced sampling for the conformational ensemble of small and mostly disordered peptides with transient α-helical structures [46, 380], which is similar to Htt17 in aqueous solution [16, 319, 459, 504]. It is also recognized as one of the best forcefield to study protein folding [28, 98, 279, 383]. Our simulations are performed in the NVT ensemble and the temperature is maintained by the Bussi–Donadio–Parrinello thermostat with a coupling constant of 0.1 ps [75]. Van der Waals and short range electrostatic interactions are cutoff at 1.0 nm. Long range electrostatics are computed using smooth Particle-Mesh Ewald [116, 150]. Bond lengths and TIP3P water geometry are respectively constrained using LINCS [184] and SETTLE [326] allowing an integration time step of 2 fs. The center-of-mass motion is removed every 20 fs. Configurations are saved every 4 ps for analysis.

We use Hamiltonian replica-exchange metadynamics (HREXMetaD) to efficiently sample the conformational ensemble and unveil the free energy landscape of Htt17 in aqueous solution. This method combines two sampling enhancing techniques : Hamiltonian replica-exchange (HREX) [256, 492] and metadynamics (MetaD) [20, 259]. MetaD introduces a history-dependent bias constructed by adding gaussians in the energy space to previously visited states along a set of specified collective variables (CVs). This increases the overall sampling at the same time as reconstructing the free energy landscape along those CVs ($\vec{S}$) as the introduced history-dependent biased potential converges to

$$V(\vec{S},t \to \infty) = -\frac{\Delta T}{T+\Delta T} F(\vec{S}) + C \tag{9.1}$$

where V($\vec{S}$,t) is the biased potential, F($\vec{S}$) is the free energy, T is the temperature, C is an irrelevant constant, and ΔT is a parameter that controls the extend of barrier heights sampled in the well-tempered flavour of MetaD [20]. It is also possible, as we do in this study, to reconstruct the free energy landscape along any omitted CV given sufficient sampling [61, 465]. The free-energy landscape is normalized so that all free energies are measured with respect to the most stable structure for each simulation, which is set at 0 kJ/mol.

For correct free energy landscape reconstruction, all slow CVs need to be considered, however the maximum number of CVs for computationally accessible is about 2-3 for MetaD which is clearly not enough to model protein folding [19, 258, 450]. One of the most efficient way to avoid this limitation is to couple MetaD with a replica exchange scheme such as Hamiltonian replica exchange (HREX) as this technique, which is widely use to simulate protein folding by its own, increases the probability of escaping free energy minima by allowing exchanges between simultaneous MD simulations at different Hamiltonians [74, 492]. Using replica exchange schemes such as HREX and parallel tempering (PT) together with MetaD allows to correctly sample other CVs not explicitly taken into account by the time-dependent biased potential as demonstrated for proteins with similar conformational ensemble to Htt17 [18, 76, 77].

For the MetaD part of our hybrid simulations, we use two CVs to bias the α-helical character

(S_α) and the radius of gyration (S_{gyr}) of the peptide :

$$S_\alpha = \sum_{i=0}^{13} \frac{1-\left(\frac{d_i}{d_0}\right)^6}{1-\left(\frac{d_i}{d_0}\right)^{12}} \tag{9.2}$$

$$S_{gyr} = \left(\frac{\sum_{i=0}^{17} |r_i - r_{\text{COM}}|^2}{\sum_{i=0}^{17} m_i} \right) \tag{9.3}$$

where the sum in S_α is over the 13 possible α-helix hydrogen bond distances d_i between main chain H_N – O couples separated by 4 residues, d_0 is 0.3 nm, the sum in S_{gyr} is over all C_α, r_i and m_i are the current C_α coordinate and mass respectively, and r_{COM} is the center-of-mass coordinate. Note that by construction $\max(S_\alpha) = 13$, but the single α-helix has $S_\alpha \sim 12.0$ as d_i is about 0.20–0.25 nm for a hydrogen bond. This choice of CVs is motivated by the fact that the Htt17 peptide in aqueous solution has an average α-helix probability of 10 to 55% according to circular dichroism (CD) spectroscopy measurements [16, 319, 459, 504], but no stable α-helix as determined by solution NMR experiments [459]. The free energy landscape of peptides with similar conformational ensemble to Htt17 were also characterized using this set of CVs [18, 76, 77]. During our simulations, a new gaussian is added to the biased potential every 4 ps with standard deviations of 0.1 and 0.01 nm along S_α and S_{gyr} respectively, and their initial height is 0.5 kJ/mol. The bias factor of the well-tempered scheme is set to 15.

The HREXMetaD simulations are performed at 303K using 16 scales for Htt17 and 24 scales for $Htt17Q_{17}$ and $Htt17Q_{17}P_{11}$ spanning 1.0 to 0.3 with intermediate scales specified by a geometric distribution as previously done [74]. Exchanges between neighboring scales are attempted every 4 ps resulting in an exchange rate of about 20–40%.

Simulated systems. All performed simulations are summarized in Table 9.I. The three investigated fragments of Huntingtin amino-terminus – Htt17, $Htt17Q_{17}$ and $Htt17Q_{17}P_{11}$ – are simulated using HREXMetaD at 303K. The P_{11} segment corresponds to the first complete proline repeat sequence of the 36-amino-acid proline-rich segment connected to Q_N in huntingtin. Addition of both Htt17 and P_{11} are sufficient to reproduce the main characteristics of the aggregation of Huntingtin's first exon [405]. The two initial states for Htt17 are a random coil structure and the solution NMR model (PDB 2LD2) determined in the presence of DPC micelles [318, 319]. In this latter state, Htt17 is an α-helix from residues 6 to 17 and disordered

for the first five residues. For Htt17Q_{17} and Htt17$Q_{17}P_{11}$, Htt17 is the NMR model, while Q_{17} and P_{11} are disordered. Random coil configurations are generated with 100 ns high temperature (600K) simulations starting from initially totally extended structures.

Peptides are solvated in a rhombic dodecahedron periodic cell and neutralized by the addition of two chloride ions. All systems are energy minimized using the conjugate gradient algorithm, and are equilibrated in the NPT ensemble at 303K for 5 ns restraining the protein backbone atoms to their initial position using a harmonic potential. All replicas are further independently equilibrated at their respective Hamiltonian scale in the NVT ensemble for 10 ns.

Our analysis is performed using in-house, GROMACS and PLUMED utilities. The secondary structure is computed using STRIDE [164] and chemical shifts using SPARTA+ [429] and Camshift [248]. All computed quantities are re-weighted to remove the bias introduced during the MetaD simulations as previously described [465] using a python implementation by Ludovico Sutto that is available to the PLUMED community. The free energies are re-weighted using a recently developed time-independent free energy estimator [465]. GROMACS utilities are used to compute the structural clusters using the gromos algorithm with a RMSD cutoff of 0.15 nm on the backbone atoms (g_cluster) [119], the solvent accessible surface area of the non-polar residues (g_sas) [146], and the occurrence of H-bonds using a cutoff of 0.35 nm on the donor-acceptor distance and of 30° on the hydrogen-donor-acceptor angle (g_hbond). Salt-bridges are considered when the distance between two oppositely charged moieties is less than 0.4 nm [21].

Errors correspond to one statistical standard deviation computed on the converge interval with 20-ns subsets.

Convergence. We assess the convergence of our simulations using three quantitative criteria as shown in Figure II.1 for Htt17, Figure II.2 for Htt17Q_{17} and Figure II.3 for Htt17$Q_{17}P_{11}$. First, we track the evolution of the global uncertainty on the free energy to identify the time at which it becomes small enough. Second, we monitor the total free energy bias added each 10 ns as a function of time to confirm that the added biases become small enough at some time. Third, we compute the two-dimensional free energy uncertainty landscape as a function of S_α and S_{gyr} on the time-interval of convergence determined from the two previous criteria to confirm that the errors are located in unimportant regions of the landscape.

We confirm that the choice of initial state does not impact our results on Htt17 as described

in details in the Supporting Material. To do so, we confirm that the convergence analysis and free energy landscape generated starting from the NMR model obtained in the presence of DPC micelles and a random coil state are very similar (compare Figure II.1 and Figure II.4). Moreover, as described in details in the Supporting Material, we ensure that our results on Htt17 are mostly independent of the sampling method by comparing HREXMetaD to PTMetaD since the latter one is most often used.

In complement, we probe the quality of the sampling of our HREX simulations by monitoring the replicas visiting the first scale that is used in our analysis as well as the secondary structure content as a function of the scaling (Figure II.5). We find great diffusion in the replica space and a more disordered peptide at larger scales.

Overall, our convergence evaluation indicates that the following analysis time intervals are suitable : 400–900 ns for Htt17_nmr (Figure II.1), 500–900 ns for Htt17Q_{17} (Figure II.2), and 500–900 ns for Htt17$Q_{17}P_{11}$ (Figure II.3). All analysis presented are thus performed on these intervals.

9.4 Results

9.4.1 Htt17

The two-dimensional free energy surface (FES) the Htt17 sequence in terms of the two biased CVs (S_α, number of helical H-bonds ; S_{gyr}, gyration radius) for the Htt17_nmr simulation is shown in Figure 9.1. It is characterized by a single large basin with the configurations below the 5 kJ/mol isoline being bound by 2 to 6 helical H-bonds and a gyration radius between 0.6 and 0.8 nm. In this region, the free energy average is 4.8 kJ/mol with the lowest energy corresponding to a S_α of 2.48 $\pm$ 0.06 and a S_{gyr} of 0.60 $\pm$ 0.01 nm. Outside this region, the energetic stability decreases gradually rising to an average of 9.1 kJ/mol as the number of helical H-bonds increases. These latter conformations are less collapsed than those below the 5 kJ/mol isoline as indicated by their larger gyration radius.

The per residue secondary structure of Htt17 is shown in Figure 9.2. We find that the first half of the peptide forms an α-helix (residue 3 to 7, $\sim$40–55%) more often than the second half ($<$35%). We also observe that residues 10 to 13 are very likely to form a turn indicating a population of two-helix bundle conformations. Overall, the peptide is mostly unstructured with only 29.3 $\pm$ 0.7% of α-helix probability, and a negligible amount of β-sheet and β-bridge

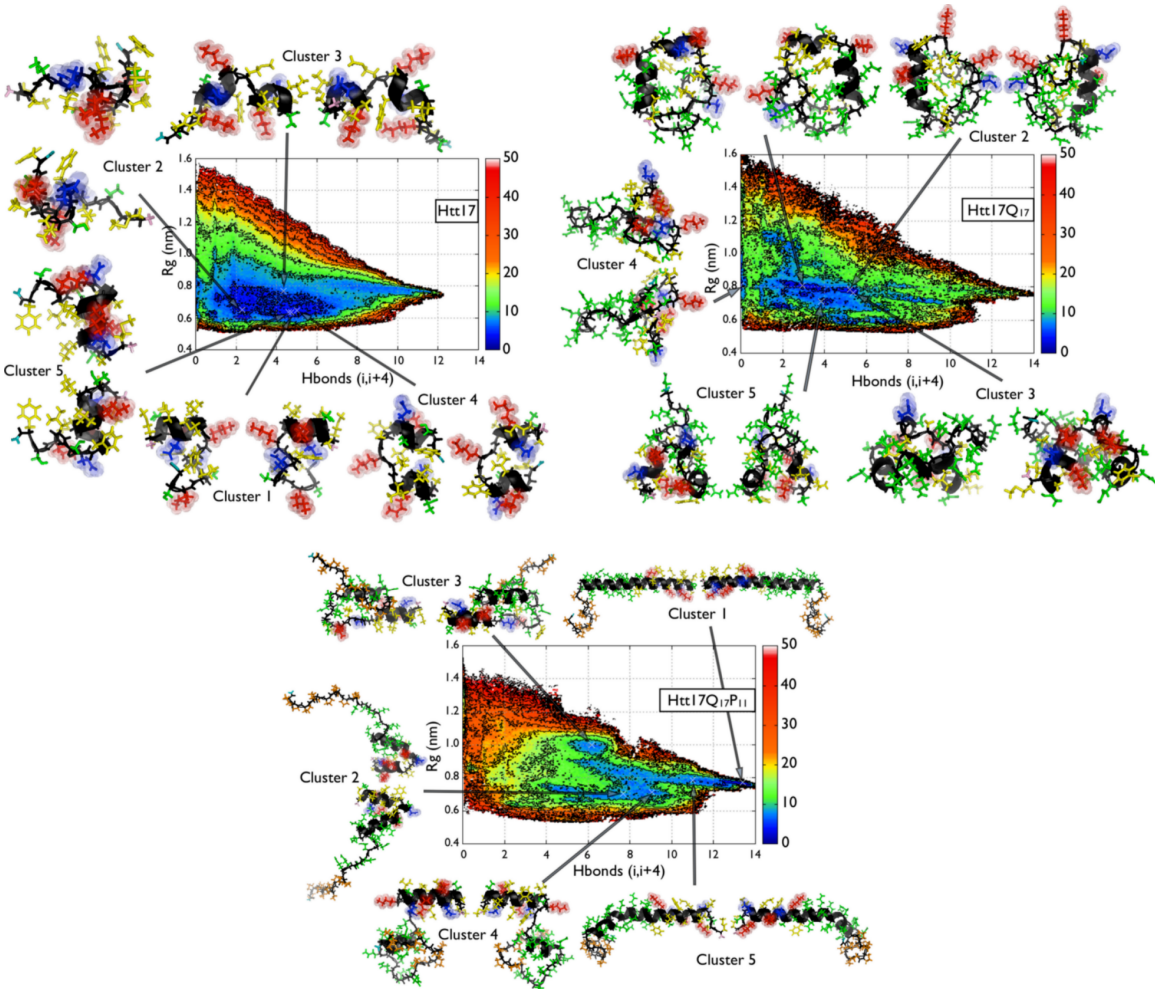


Figure 9.1 – The free energy landscapes of Htt17 (Htt17_nmr) (shown in the top left panel), Htt17Q_{17} (shown in the top right panel) and Htt17$Q_{17}P_{11}$ (shown in the bottom panel). The horizontal and vertical axes respectively represent the number of helical H-bonds (S_α) and the gyration radius (S_{gyr}). The number of helical H-bonds is computed on the first 13 residues for Htt17 and the first 17 residues for Htt17Q_{17} and Htt17$Q_{17}P_{11}$ to include possible H-bond formation with the Q_{17} domain. The energy isolines are drawn every 5 kJ/mol. The uncertainty on the free energy landscapes of Htt17, Htt17Q_{17} and Htt17$Q_{17}P_{11}$ are respectively shown in Figures II.1, II.2 and II.3. The uncertainty is always smaller than 1 kJ/mol on the relevant parts of the landscapes. In addition, the cluster analysis of the most representative conformations populating the FES below 5 kJ/mol and below 8 kJ/mol for Htt17$Q_{17}P_{11}$. The negatively charged, positively charged, non polar and polar residues of Htt17 are shown in blue, red, yellow and green. The Q_{17} and the P_{11} segments are respectively coloured in green and orange. The backbone is coloured in black, the amino-terminus in pink and the carboxy-terminus in teal.

totalizing 1.2 $\pm$ 0.2%. These propensities are in agreement with measurements from several circular dichroism (CD) experiments that indicate a helical propensity between 10–55% [16,

319, 459, 504].

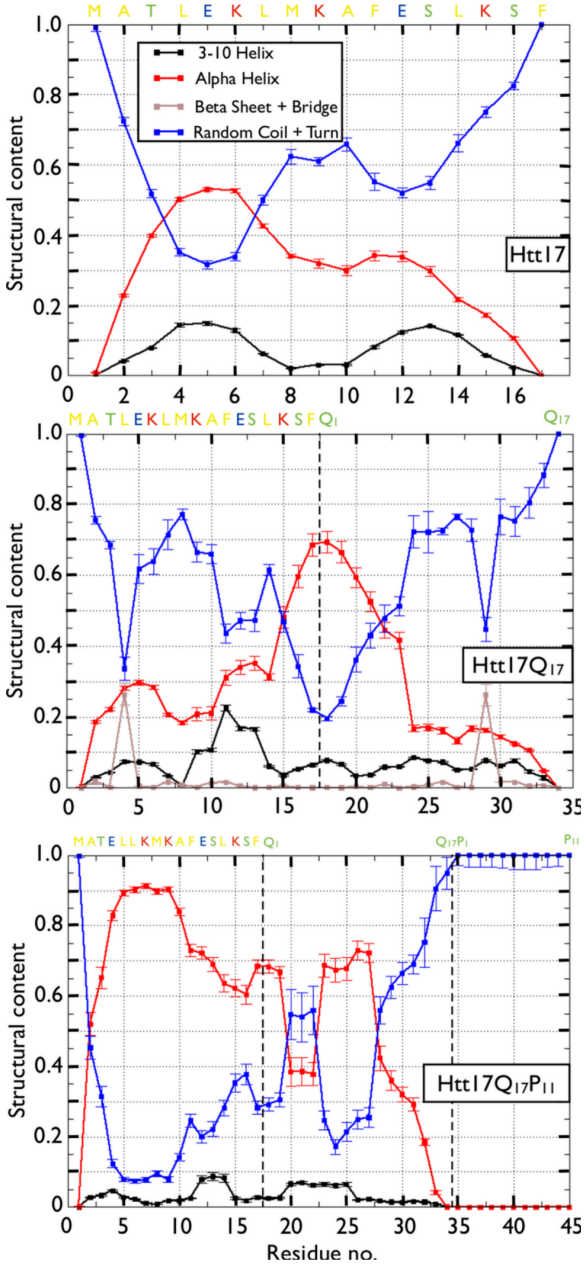


Figure 9.2 – From top to bottom, the per residue secondary structures of Htt17 (Htt17_nmr), $Htt17Q_{17}$ and $Htt17Q_{17}P_{11}$. The probability of α-helix, 3-10 helix, β-bridge and β-strand, and turn/coil are respectively shown in red, black, brown and blue. The vertical black dotted line for $Htt17Q_{17}$ and $Htt17Q_{17}P_{11}$ indicates the end of the Htt17 segment. The average and standard deviation are computed on the converged interval using 20-ns time windows.

We also evaluate the presence of structural elements using the secondary H_α chemical shift

and NOE signal sampled in our simulations (Figure 9.3) as to compare with the solution NMR experiment of Wetzel and co-workers [459]. The secondary chemical shifts are in very good agreement with the NMR measurements ; except for residues Lys9, Ala10 and Ser16 that are more extended in our simulations. For the most part, the secondary chemical shifts are small and negative indicating weak helical features. The NOEs signal of the peptide in our simulation, for its part, indicate a mostly unstructured state as there are high H^N(i)-H^N(i+1) and low H^α(i)-H^N(i+2) signals, in agreement with the NMR experiment [459]. However, small H^α(i)-H^N(i+3) NOE intensities suggest the presence of a small population of more helical structures (see Supporting Material for an in-depth comparison). Therefore, the presence in our simulations of a global α-helix average of about 30% without any individual residue showing more than 55% seems to slightly overestimate the α-helical propensity but is overall compatible with the NMR data.

The main configurations sampled by Htt17 in basins below the 5 kJ/mol isoline are depicted by their clusters' center in Figure 9.1. In line with our previous analysis, the first residues of the Htt17 peptide have a greater tendency to be helical, while the last residues are mostly unstructured and the central part of Htt17 (residues 10 to 13) forms a turn bringing the amino- and the carboxy-terminus close to each other (see clusters 1, 3 and 4).

To get a more quantitive view on the tertiary structure, we computed an additional two-dimensional FES as a function of the solvent accessible surface area (SASA) of Htt17's non-polar residues and the number of helical H-bonds (Figure II.6). We see that the conformations sampled are mostly characterized by a SASA of 3 to 5 nm^2 indicating that the non-polar residues are mostly accessible to the solvent. These results are very similar for a disordered Htt17 also characterized by a 3 to 5 nm^2 SASA for an average of 4.10 ± 0.01 nm^2. A closer look at the main clusters (Figure 9.1) reveals indeed that the non-polar residues are mainly accessible to the solvent.

The contact map between each residue is indicative of the mostly flexible and disordered tertiary structure of Htt17 as most contacts are between neighboring residues (Figure II.7). We observe, nevertheless the presence of three electrostatic contacts : Glu5–Lys9, Glu12–Lys9 and Glu12–Lys15 with a probability of 50.3%, 42.3% and 64.0%, respectively. The formation of a stable salt-bridge occurs less often however with probability of $4.2 \pm 0.1\%$, $11.0 \pm 1.0\%$ and $12.3 \pm 0.3\%$, respectively. We also note a long range non-polar contact between Met8 and Phe17

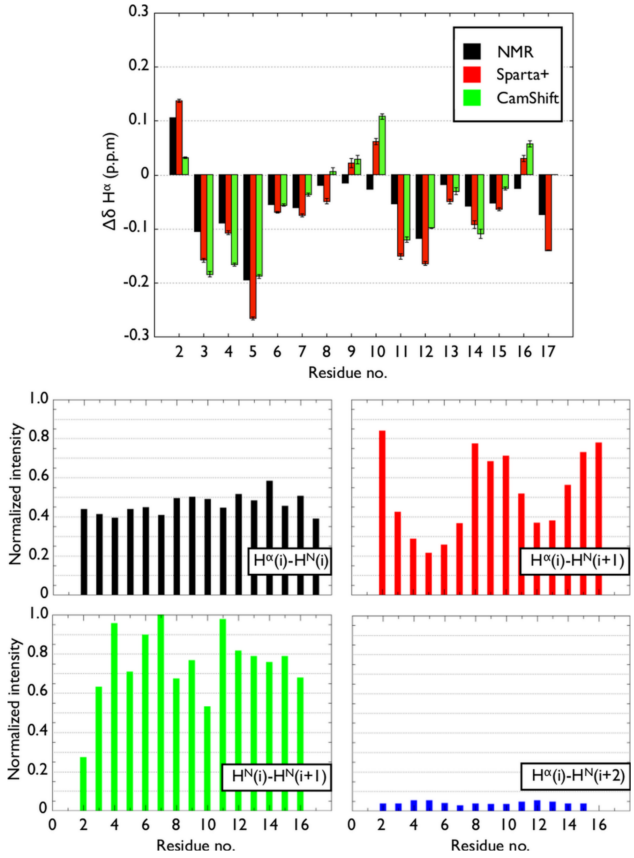


Figure 9.3 – **(A)** The H^{α} secondary chemical shift per residue computed with SPARTA+ (red) and CamShift (green) are compared to the NMR measurements (black) on Htt17 in aqueous solution [459]. The secondary shifts are obtained by subtracting the corrected coil value specific to each amino acid type to their H^{α} chemical shift [415]. **(B)** The computed intensities of the interproton NOEs for all residues between the H^{α} of residue i and the H^{N} of residues i, i+1 and i+2, as well as between the H^{N} of residues i and i+1 are compared to the equivalent NMR measurements on Htt17 in aqueous solution [459].

in 24.1% of the sampled conformations that seems to be involved in the formation of the turn between residues 10 and 13 as well as in the destabilization of the secondary structure in the second half of the peptide (see clusters 1 and 4 of Figure 9.1).

9.4.2 Htt17Q_{17}

We now investigate the changes induced on the FES of Htt17 due to the addition of the Q_{17} segment at its carboxy-terminus. The two-dimensional FES of Htt17Q_{17} in terms of S_{α} and S_{gyr} – computed only on the Htt17 region – is shown in Figure 9.1. In the largest basin, we observe three distinct minima characterized by a similar gyration radius around 0.8 nm, but with

a different number of helical H-bonds (either 0.0 or between 2.0 and 6.0). The lowest point of the FES has a gyration radius of 0.69 $\pm$ 0.01 nm and 1.9 $\pm$ 0.1 helical H-bonds. Addition of the Q_{17} region modifies key features of the FES of Htt17 as it becomes overall more extended (more configurations having a larger gyration radius) and less structured (fewer configurations having a large number of helical H-bonds) as shown in Figure 9.1. More precisely, the free energy of the conformations with a large number of helical H-bonds significantly increases from an average of 9.1 kJ/mol for Htt17 alone to 14.6 kJ/mol for Htt17Q_{17}.

In terms of secondary structure, we observe a significant loss of helical propensity for residues 2 to 10, while that of the remaining residues in Htt17 greatly increases upon addition of the Q_{17} segment (Figure 9.2). Even though the Htt17 segment in Htt17Q_{17} has an overall α-helix probability that is unchanged with respect to Htt17 alone (30.4 $\pm$ 1.4 % vs. 29.3 $\pm$ 0.7%, respectively), the per residue probability is very different : there is a significant amino-to-carboxy-terminus shift of the helical probability that is directly due to the presence of the Q_{17} as the α-helix in Htt17 continues up to the first six glutamines. The remaining part of the Q_{17} is however mostly disordered. We also note the presence of a turn between Glu5 and Leu7 ($\sim$40-45%) as well as a β-bridge between Leu4 and a glutamine (26 $\pm$ 3%). With the exception of Leu4, the overall amount of β-sheet and β-bridge is still negligible. Overall, the global α-helix probability of Htt17Q_{17} is 29.1 $\pm$ 1.4%.

A cluster analysis of all sampled structures characterized by a free energy of less than 5 kJ/mol deepens the atomistic insights (Figure 9.1). The first five clusters can be classified in three main categories : (1) no helical H-bond in Htt17 and a fully disordered Q_{17} as shown by clusters 2, (2) two small α-helix fragment at both end of Htt17 and a disordered Q_{17} as shown by cluster 4 and 5 and (3) an α-helix spanning the last residues of Htt17 and the first glutamines of Q_{17} as shown by clusters 1, 3. In these clusters, the Htt17 segment adopts a u-shaped topology with a turn around Leu7 bringing its non-polar residues close together as shown by Htt17Q_{17}'s contacts map (Figure II.7). More precisely, we identify long-range non-polar contacts between Leu4–Phe17 (20.8%), Met8–Phe17 (20.4%) and Phe11–Phe17 (29.1%) that form a non-polar cluster, which is further isolated from the solvent by the aliphatic carbons of the glutamines (for example, see clusters 1 and 3 of Figure 9.1). As in Htt17 alone, we observe electrostatic contacts/saltbridge between Glu5–Lys9 (44.5% / 6.9 *om* 0.2%), Glu12–Lys9 (51.1% / 23 $\pm$ 2%) and Glu12–Lys15 (48.9% / 12.9$\pm$ 0.2%). The charged residues Glu5/Lys6/Lys9/Glu12/Lys15 of

the Htt17 domain also interact with the Q_{17} domain with a probability of 54.6%, 50.8%, 48.2%, 40.7%, 83.9% respectively. For its part, the glutamines aliphatic carbons of the Q_{17} segment interacts a lot with the non-polar residues of the Htt17 segment. As a consequence, the resulting non-polar SASA decreases to as low as 2 nm^2 in contrast to Htt17 alone that as a non-polar SASA between 3 and 5 nm^2 (Figure II.6). We also note the presence of several main-chain/side-chain and main-chain/main-chain interactions between Htt17 and Q_{17} ($\sim$ 30%, Figure II.7).

The FES of the Q_{17} region in terms of the S_α and S_{gyr} CVs shows that it is mostly disordered and collapsed (Figure II.8). A cluster analysis on the configurations with less than 5 kJ/mol reveals that first glutamines are α-helical, while the remaining part of the Q_{17} region is disordered independently of Htt17's structure, which is either an α-helix from residues Ser13 to Phe17 (clusters 1, 4 and 5) or from residues Ala2 to Met8 (clusters 2). At high population of helical H-bonds, we note the presence of a very narrow minimum characterized by an almost fully α-helical Q_{17} with Htt17 adopting the same structure up to residue Glu12.

9.4.3 Htt17$Q_{17}P_{11}$

We finally probe the effects of the addition of Q_{17} and P_{11} segments on the FES of Htt17. The two-dimensional FES of Htt17Q_{17} in terms of S_α and S_{gyr} – computed only on the Htt17 region – is shown in Figure 9.1. The resulting FES unveils a striking shift toward the α-helix as the FES is characterized by a single minimum with a number of helical H-bonds between 12 and 14 and a gyration radius between 0.75 and 0.85 nm. The lowest point on the FES has 13.3 $\pm$ 0.1 helical H-bonds and a gyration radius of 0.76 $\pm$ 0.01 nm. Most of the structures sampled by Htt17 alone or Htt17Q_{17} are therefore less stable upon the addition of P_{11}.

Analysis of the secondary structure indicates that the Htt17 region adopts α-helical conformations 70.9 $\pm$ 1.6% of the time, a drastic increase compared to both Htt17 (29.3 $\pm$ 0.7%) and Htt17Q_{17} (30.4 $\pm$ 1.4 %) in solution (Figure 9.2). The probability is notably high for residues 5 to 9 that have 90%. For its part, the Q_{17} domain has 44.8 $\pm$ 2.5% of α-helical content with the residues near Htt17 (first, second and sixth to tenth glutamines) having the largest probability, while glutamines 3 to 5 and those near P_{11} mostly form turn/coil structures. Finally, the P_{11} domain almost exclusively forms a PPII-helix characterized by Φ and Ψ dihedral angles respectively near of -75° and 150° according to an analysis of the Ramachandran plot for every proline (data not shown).

The addition of P_{11} causes drastic changes in structure as it doubles the helical content of both Htt17 and the Q_{17} domain. We further quantify its effect from a cluster analysis on the structures characterized by a free-energy less than 8 kJ/mol. In contrast to the important structural diversity of Htt17 and Htt17Q_{17} in solution, Htt17 within Htt17$Q_{17}P_{11}$ has a strong tendency of forming an α-helix as the first cluster is composed of more than 25% of the sampled structures. The depicted conformation for the first cluster shows the Q_{17} as a fully formed α-helix and the P_{11} region extends away from Htt17 and Q_{17}. A more representative analysis of the structural ensemble of Q_{17} is given in the next paragraph. We note a clear separation between the polar and non-polar residues of Htt17. The formers are interacting mostly with Q_{17} as shown on Htt17$Q_{17}P_{11}$ contacts map (Figure II.7). More specifically, there are contacts between Glu5/Lys6/Lys9/Glu12/Lys15 of the Htt17 domain and the glutamines with a probability of 25.0%, 21.0%, 61.1%, 57.6% and 85.7%, respectively. Salt-bridges are also present between Lys9–Glu12 (45 $\pm$ 4%) and Lys15–Glu12 (14 $\pm$ 2%) stabilizing Htt17 conformation. For its part, the P_{11} domain interact mostly with residues surrounding Lys9 (33.5%), Ser13 (50.0%) and Leu14 (24.0%) and Phe17 (48.8%) via mostly their side-chain. As for the non-polar residues in Htt17, they are all located on the same side of the peptide and fully accessible to the solvent as shown by a drastic increase of the non-polar SASA when compare to the Htt17 and Htt17Q_{17} peptides (Figure II.6).

The FES of Q_{17} as a function of our two CVs (S_α and S_{gyr}) unveils three minima with distinct number of hydrogen bonds (either 0, 2.5 and 7.5) as shown in Figure II.8. A cluster analysis on the structures found inside those region (below 4 kJ/mol) shows that the most important cluster of the Q_{17} domain has an important helical propensity up to the tenth glutamine and that the remaining glutamines are mostly unstructured. The other clusters depict the Q_{17} domain as fully unstructured independently of Htt17's structure that is either a fully formed α-helix (clusters 2, 4 and 5) or mostly unstructured (cluster 3).

9.5 Discussion

Numerous experiments indicate that the Huntingtin amino-terminus is crucial for its biological functions. More specifically, the first 17-amino-acid segment (Htt17), which is right before the amyloidogenic polyglutamine segment (Q_N), is directly involved in the membrane interactions and aggregation of Huntingtin. In this study, we quantify the conformational ensemble of three

fragments of Huntingtin amino-terminus – Htt17, Htt17Q_{17} and Htt17$Q_{17}P_{11}$ – in terms of free energy surfaces, secondary structures, contact maps and clusters. Our results demonstrate the effects of Q_{17} and P_{11} on the conformational ensemble of Htt17 and, taken together with other studies, they provide insights on motifs at the origin of Htt17's membrane-binding and oligomerization.

9.5.1 Htt17 samples a wide variety of coil/helix structures

Experiments indicate that Htt17 has a helical population of about 10–55% in aqueous solution using circular dichroism (CD) [16, 319, 459, 504], but no stable structural motif according to solution NMR measurements [459]. Recently results from ion mobility spectrometry-mass-spectrometry (IMS-MS) coupled to molecular dynamics simulations suggest that Htt17 populates two kinds of helical monomer with an α-helix from the amino-terminus to residue Lys9–Ala10 : (i) a compact structure characterized by an unstructured region between residues Phe11 and Phe17 that turn back on itself and brings the amino- and carboxy-terminus closer to each other and (ii) an extended structure characterized by a 3-10 helix spanning residues Ala10 to Glu12 and where the carboxy-terminus region is extended away from Htt17 [13]. Taken together, these observations suggest that the structural ensemble of Htt17 consists of a wide variety of flexible helix/coil conformations.

Previous simulations on Htt17 suggest such a conformational ensemble [233, 401, 504]. More precisely, simulated tempering simulations with the AMBER03 forcefield and explicit solvent (TIP3P) show that the conformational ensemble of Htt17 contains about 70% of diverse two-helix bundles with a loop around Ala10, while the rest of the ensemble populates a single straight helix or disordered configurations [233]. In this work, residues 3 to 6 have the highest α-helix propensity and that the sampled configurations are mostly stabilized either by charged interactions or sequestration of the non-polar residues. Other simulations were performed using Monte Carlo with the ABSINTH implicit solvent forcefield showing that Htt17 has an α-helix probability of 34% and that it is mostly collapsed upon itself to sequester its non-polar residues [504]. Bias-exchange metadynamics, for its part, unveiled the free energy landscape of Htt17 using explicit solvent, all-atom simulations (AMBER99/TIP3P) [401]. In this work, Carloni *et al.* observe that the free energy landscape using 6 collective variables (CVs) is mainly made of four basins and that the transitions from one basin to the others occur on the microse-

cond timescale. The resulting conformational ensemble of Htt17 is largely disordered (75%) and helical (25%) with a global α-helix probability of 29% notably for residues 1 to 7. They also note that the disordered configurations of the largest basin have their non-polar residues largely accessible to the solvent.

This is in line with the trend depicted by our simulations in terms of the free energy landscape (Figure 9.1) and the secondary structure propensity (Figure 9.2). Our predicted secondary structure is characterized by a global α-helix probability of 29.3 $\pm$ 0.7% that is similar to the values obtained in the aforementioned simulations – 43% [233], 34% [504] and 29% [401] – and CD experiments – 10% [319], 34% [504], 45% [16] and 55% [459] – on Htt17 in aqueous solution. Our results also indicate that residues 3 to 5 have the highest helix propensity (Figure 9.2, about 40–50%) in agreement with other simulation protocols [233, 401]. Finally, we find that Htt17 forms various two-helix, single helix, helix/coil and coil conformations as previously observed [13, 233, 401]. The probability of structured conformations is however lower in our simulations than in Ref [233], which might be due to AMBER03 slightly overstabilizing helical structures in helix/coil peptides when compared to the AMBER99sb*-ILDN force field as indicated by other studies [46, 380].

In terms of tertiary structure, our simulations indicate that Htt17's non-polar residues are mostly accessible to the solvent (Figure II.6) in agreement with previous bias-exchange metadynamics simulations [401]. In addition, mainly short-range contacts between neighbouring residues are populated in Htt17. Still, a non-polar contact between Met8 and Phe17 occurs in 24.1% of the sampled structures. It could be crucial in the formation of the turn between residue Ala10 and Ser13 therefore leading to the destabilization of the second half of Htt17.

Finally, we provide a detailed analysis indicating that the structural ensemble sampled in our simulations is consistent with the only solution NMR experiment done on Htt17 in an aqueous environment [459] in terms of secondary H_α chemical shifts and interproton NOE distances (Figure 9.3 and Supporting Material). We are thus confident that our results yields relevant insights on the structural ensemble of Htt17.

9.5.2 Addition of Q_{17} reduces Htt17's non-polar residues accessibility to the solvent

Fluorescence-based resonance energy transfer (FRET) experiments indicate that Htt17 is in a collapsed state and that it becomes more extended upon addition of the polyglutamine seg-

ment [459]. CD spectra suggest an increase of helix propensity with the polyglutamine length, but it is unknown if it is due to the Q_N or Htt17 segments [209]. Data from X-ray crystallography on a chimeric protein containing Huntingtin exon 1 supplement this by indicating that an α-helix in Htt17 can extend to the Q_N region [239]. In both studies, absence of β-sheet is observed. For its part, Monte-Carlo simulations using the implicit solvent ABSINTH potential show that addition of the Q_N domain disorders Htt17 in a length-dependent manner, while the Q_N segment itself remains disordered [504]. Pappu *et al.* also find that the non-polar residues of Htt17 lie in interdomain interface between Htt17 and Q_N.

Our results complement these experiments by showing that the Q_N region in Htt17Q_{17} is mostly disordered, but that it can sample α-helices with a probability of 27.8 $\pm$ 1.3%, particularly for the first glutamines (Figure 9.2). The Q_N region also induces an amino-to-carboxy-terminus shift of the helical probability in the Htt17 region but leaves unchanged its global α-helical probability from 29.3 $\pm$ 0.7% to 30.4 $\pm$ 1.4%. In line with experimental results, we find negligible amount of β structures for both the Htt17 and Q_N regions at the monomer level. We also find that the aliphatic carbons of the Q_N domain interact directly with the non-polar residues of Htt17 dramatically reducing their solvent accessibility (Figures II.6 and II.7), in agreement with previous simulations [504].

Overall, we observe that the Q_N region modifies the conformational ensemble of Htt17 already at the monomer level, which could have a direct impact on its aggregation and membrane-binding affinities as discussed next. This indicates that not only does Htt17 influence Q_N as previously determined experimentally, but that the opposite also occurs and that the interplay between Htt17 and Q_N might be more complex than previously thought.

9.5.3 Htt17 is more structured upon addition of $\mathbf{Q_{17}P_{11}}$

Circular dichroism (CD) measurements show that the addition of a P_{10} domain to Htt17Q_{37} reduces the the α-helix probability from more than 50% to around 30% [209]. However, CD is unable to tell the localization of these structural changes. Using HPLC sedimentation assay, Wetzel et al. also reported that the aggregation of Htt17Q_{35} is quicker than for Htt17$Q_{37}P_{10}$, although the latter is still much faster than a Q_N domain of similar length alone. ThT fluorescence kinetic profile monitoring the growth of the fibril showed mostly no difference between Htt17Q_{30} and Htt17Q_{30}C38, where C38 is the full-length proline-rich region starting with P_{11},

indicating that the aggregation mechanism is dominated by Htt17 [504]. It is also found that C38 acts as a solubilizing module that weakens the driving force toward the formation of insoluble aggregates. X-ray crystallography on a chimeric protein containing Huntingtin exon 1 suggests that the Htt17 can populate α-helix configurations, while the Q_N region is mostly unstructured except for the first glutamines that can populate an α-helix [239]. The first prolines in the proline-rich region, for their part, are characterized by a PPII-helix.

Only one set of simulations has been performed on Htt17$Q_N P_{11}$ to our knowledge [137]. These all-atom replica exchange molecular dynamics simulations with the FF03 force field and implicit solvent suggested that both Htt17 and Q_N adopt mostly α-helical conformations, while the P_{11} forms a PPII-helix. In these simulations, the α-helical content is especially large between residues 4 and 17 of Htt17, and the P_{11} region lies anti-parallel to the Htt17 region when there are 17 glutamines in Q_N, but not when there are 55 glutamines (above than the pathological threshold of 36 repeats).

In our simulation, the P_{11} domain stabilize Htt17 as an almost fully formed α-helix with more than 70% of helical content. The Q_{17} domain adopts an α-helix conformation 44.8 $\pm$ 2.5% of the time. Our results differs from the secondary structure signal from CD [209] and are surprising overall. Indeed, the increase of the non-polar SASA and the decreased number of contacts between Htt17 and Q_{17} are two strong destabilizing factors present in our simulations. The difference with experiment is perhaps due to the length of the Q_N domain used ; longer Q_N (as in the CD experiment) might mitigate the stabilizing effects of P_{11}. An other simulation protocol show a similar α-helical population in Htt17 in the context of Htt17$Q_{17}P_{11}$, but significantly more α-helix in Q_{17} [137]. Ultimately, other simulation and experimental protocol will be needed to unveil the origin of this dissimilarity.

9.5.4 Motifs relevant to membrane-binding and oligomerization

Htt17 is crucial to the localization of Huntingtin in the cell [4, 14, 16, 175, 302, 398, 443, 463, 536, 537] and adopts an α-helical conformation in the presence of micelles, vesicles and phospholipid membranes as shown by CD spectroscopy [16, 319], solution NMR [319], solid-state NMR [318] and Hamiltonian replica-exchange simulations [106]. Solid-state NMR and Hamiltonian replica-exchange simulations also indicate that the amphipatic plane of Htt17 is aligned parallel to the membrane surface with its non-polar residues facing the hydrophobic

core of the membrane. The presence of α-helical structures in Htt17 prior to membrane-binding has been shown to ease its insertion in the membrane [111].

Htt17 also drastically modifies the aggregation of the amyloidogenic Q_N segment. Three main models have been proposed to describe Htt17's role in the aggregation mechanism of Huntingtin : (i) the formation of tetrameric α-helical bundles of Htt17 that increases the local concentration of Q_N favoring the nucleation of β-sheeted structures in the latter region [209], (ii) the reduction in the entanglement of Q_N destabilizing the non-fibrillar aggregates [114], and (iii) the direct interaction between Htt17 and Q_N favoring formation of extended structures in the latter region [457]. Solid-state NMR indicates, for its part, that the core of a $Htt17Q_{30}P_{10}K_2$ amyloid fibril is formed by Q_N, while Htt17 and P_{10} respectively form an α-helix and a PPII-helix [187, 434].

The first aggregation model, more specifically, is based on sedimentation velocity experiments that indicate that Htt17 and $Htt17Q_{10}K_2$ are mostly monomeric in solution with low level of compact oligomers that correspond to, in decreasing population, tetramer, octomer, dodecamer and so on [209]. The aggregation-enhancing property of Htt17 with respect to Q_N alone can be then explained by the formation of reversible α-helical tetrameric bundles via Htt17. Namely, these tetramers assemble into higher order oligomers that increase the local concentration of the amyloidogenic Q_N segment easing the nucleation of the β-sheeted structures necessary to the formation of amyloid fibrils.

These previous investigations indicate common motifs in Htt17 – the formation of helical structures and the sequestration of its non-polar residues – that are of fundamental to both its aggregation and membrane-binding enhancing properties. We now discuss how our observations on the Htt17, $Htt17Q_{17}$ and $Htt17Q_{17}P_{11}$ monomers are related to these models.

We observe in our simulations the presence of α-helical conformations ($29.3 \pm 0.7\%$) in Htt17. We also quantify, more specifically, the presence of highly α-helical structures in Htt17 by comparing the sampled ensemble in aqueous solution to its membrane-bound state. We compute the RMSD with respect to the α-helical structure sampled in a POPC bilayer [106] and reconstruct the two-dimensional FES of Htt17 in terms of this RMSD and the number of helical H-bonds (Figure 9.4). We observe a broad basin between 1.0 and 7.5 helical H-bonds and 0.3 and 0.7 nm RMSD, corresponding to structures that are different from their membrane counterpart. The free energy gradually rises as the structural similarity to the membrane state increases

indicating that a highly α-helical structure similar to the membrane-bound state is not stable in aqueous solution.

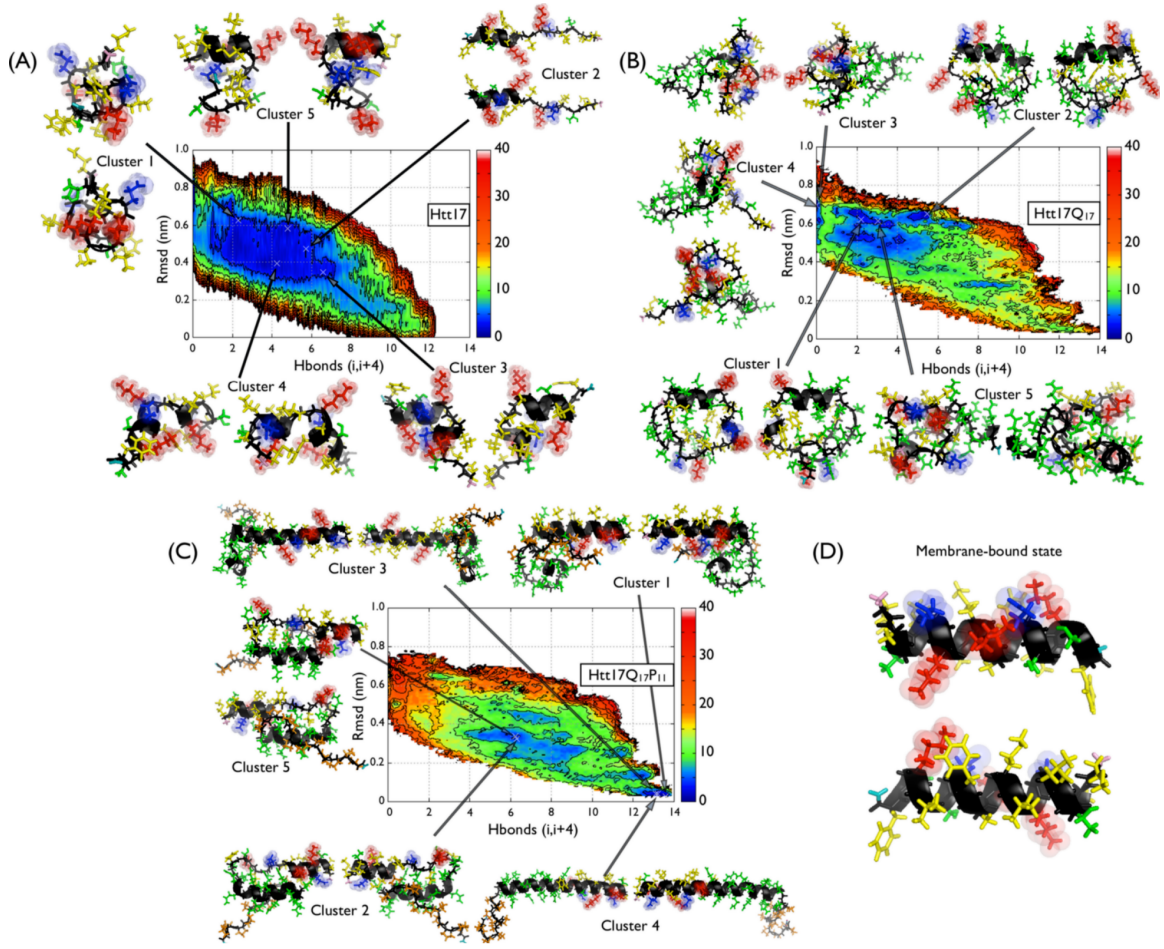


Figure 9.4 – Comparison of the conformational ensemble of Htt17 to the membrane-bound state [106]. The FES of the Htt17 segment as a function of the backbone RMSD measured between Htt17 in solution and the membrane-bound state (vertical axis) and the number of helical H-bonds (S_α, horizontal axis) is shown along with the representative principal conformation clusters for **(A)** Htt17, **(B)** Htt17Q_{17} et **(C)** Htt17$Q_{17}P_{11}$. The membrane-bound state is depicted in **(D)** The negatively charged, positively charged, non polar and polar residues of Htt17 are shown in blue, red, yellow and green. The Q_{17} and the P_{11} segments are respectively coloured in green and orange. The backbone is coloured in black, the amino-terminus in pink and the carboxy-terminus in teal. Energy isolines are drawn every 4 kJ/mol.

We note, nonetheless, that some configurations in the basin below the 4 kJ/mol isoline possess a motif that could initiate the membrane-binding and the formation of the tetrameric α-helical bundle : the first residues of Htt17, particularly between residues Thr3 to Lys6, can form an α-helix ($\sim$50%, Figure 9.2). The presence of such motif has also been observed in other

simulations [401]. Moreover, our result show that the non-polar residues of Htt17 are mostly accessible to the solvent (Figure II.6), particularly Met8, Phe11 and Phe17, which could promote the association of Htt17 with other Htt17 segments and its anchoring to a phospholipid membrane.

Upon addition of Q_{17}, the population of a highly helical Htt17 segment is significantly reduced (Figure 9.2), moving away from the membrane-bound state (Figure 9.4). This is explained by the α-helix motif between residues 3 and 7 being less stable than for Htt17 alone even though a new α-helical motif starting at residue 15 in Htt17 and extending in the first glutamines of Q_{17} is formed. This shift of the position of the α-helical motif could result in a shift of Htt17Q_N's primary interaction site with phospholipid membranes and nucleation site for the tetrameric bundles formation during oligomerization.

We also observe that the non-polar residues of Htt17 are globally less exposed due to interactions with the aliphatic carbons of the glutamines (Figures II.6 and II.7). Together with the α-helical shift, this could affect the aggregation and membrane-binding pathways as the Q_N region needs to move away from the Htt17 region to free the non-polar residues for these events to proceed. This might be one of the rate limiting steps for the tetrameric bundle formation as intrapeptide Htt17–Q_N interactions needs to be dominated by interpeptide interactions between Htt17–Q_N or Q_N–Q_N during dimerization. Previous simulations indeed indicate that the Q_N–Q_N interaction dominate in the dimer [504].

Our results also suggest that the stability of Htt17 in a fully formed α-helix state drastically increases with the addition of the $Q_{17}P_{11}$ domain (Figures 9.1 and 9.2). This results in an important population of membrane-bound like states characterized by a RMSD below 0.1 nm (Figure 9.4) and a high solvent accessibility for the non-polar residues of Htt17 (Figure II.6). Both the α-helical character and the non-polar residues accessibility of Htt17 due to the combined addition of Q_N and P_{11} could promote membrane-binding, as observed experimentally [72] and numerically [335], as well as the formation of Htt17 tetrameric bundle.

A priori, this observation from our simulation on the P_{11} role seems to contradict previous experiments. In some studies, the addition of a P_{10} domain decreases the rate of formation and the stability of amyloid-like aggregates leaving the nucleation mechanism unchanged compared to Q_N domain alone [51]. P_{10} would therefore stabilize conformations incompatible with aggregation. Other studies show, however, that the proline-rich segment C38 – starting with P_{11} and

located at the carboxy-terminus of Q_N – increases the overall solubility of Htt17Q_NC38, weakening the driving force toward the formation of insoluble aggregates, but preserving a rate of fibril formation similar to that of Htt17Q_N [114]. Consequently, the formation of structural features that could favor the aggregation of Htt17 in our simulations might not be sufficient to enhance the overall oligomerization. The slowing effect of P_{11} could then occur latter in the aggregation process or be caused by another phenomena such as induced structural changes in the Q_N domain. In that regards, our simulations on the monomer indicate that the structural flexibility of Q_N is reduced by the incorporation of P_{11} (Figure II.8)

Other models have been proposed for huntingtin aggregation. In the first one, both Htt17 and C38 – the proline-rich segment starting with P_{11} and located at the carboxy-terminus of Q_N – modulate the aggregation of Q_N by controlling the intrinsic structural heterogeneity of this amyloidogenic segment [114]. Fibrillation is promoted by Htt17 destabilizing the intermediate non-fibrillar structures and P_{11} destabilizing the intermediate insoluble aggregates. The role for Htt17 was unveiled using Monte-Carlo [504] and mesoscopic [403] simulations that investigated respectively the dimerization and large-scale aggregation of Htt17Q_N. Tthe Htt17 segment would then reduce the entanglement within the Q_N segment and introduces a barrier to intermolecular associations that brings the formation of small spherical structures (soluble oligomers) and large linear aggregates (insoluble fibrils) on similar timescale. In the second model, the amyloidogenic Q_N segment would interact directly with Htt17 to promote fibrillation through the formation of extended motifs in Q_N [457].

The relations between our simulations and these models remain, nevertheless, more limited as we have focused our investigation on the Htt17 segment and do not have a simulation on the Q_N segment alone. We observe, nonetheless, that Q_{17} in the presence of Htt17 adopts a variety of structures that are mostly disordered (Figure II.8). It also interacts directly with Htt17 (Figure II.7) as previously suggested experimentally [457]. Addition of P_{11} leads to a more compact Q_{17} region that interacts much more with itself and that this could reduce entanglement during oligomerization.

9.6 Conclusion

We studied three fragments amino-terminus of Huntingtin – Htt17, Htt17Q_{17} and Htt17$Q_{17}P_{11}$ – with special consideration to the first 17-amino-acids segment (Htt17) that is crucial for its oligo-

merization and membrane binding. We applied a novel combination of two sampling enhancing techniques – Hamiltonian replica-exchange and well-tempered metadynamics (HREXMetaD) – to have a thorough understanding of the modifications on Htt17's structural ensemble due to the addition of the amyloidogenic Q_N segment and the polyproline segment (P_{11}).

We find that the structural ensemble of Htt17 is characterized by a wide variety of helix/coil conformations. The addition of the Q_{17} domain results in an amino-to-carboxy-terminus shift of the helical content and it decreases the solvent accessibility of Htt17's non-polar residues by interacting directly with it. The addition of both Q_{17} and P_{11} drastically changes the structural ensemble of Htt17 towards more structured conformations with more exposed non-polar surfaces.

Careful comparison with experimental aggregation and membrane-binding models reveals that Htt17 possesses crucial features essential to these processes whether it is combined with Q_{17}, $Q_{17}P_{11}$ or alone. We find that the position and the type of motifs are very different depending on the adjacent sequences to Htt17 showing that all these neighboring regions strongly impact each other already at the monomer level.

Our results also provide as a strong basis for further study of more complex situations such as Htt17Q_NP_{11} oligomerization and membrane-binding using a similar simulation protocol (HREXMetaD). We find this novel protocol to offer good sampling at a moderate computational cost and to scale very well with the number of particles as it is essentially size independent.

9.7 Supporting Material

In Annexe II.

9.8 Author contributions

V.B. and S.C. designed the research ; V.B. and S.C. performed the research ; V.B. and S.C. contributed analytic tools ; V.B, S.C. and N.M. analyzed the data ; V.B. S.C. N.M. wrote the paper.

9.9 Acknowledgments

The authors thank the PLUMED community – particularly Giovanni Bussi, Pratyush Tiwary and Ludovico Sutto – for helpful discussions and advices. They also thank Ronald Wetzel and In-Ja

Byeon for kindly providing their NMR data and for their insight on our results. Computations were made on the supercomputers Briarée from Université de Montréal and Mammouth from Université de Sherbrooke, managed by Calcul Québec and Compute Canada. The operation of these supercomputers is funded by the Canada Foundation for Innovation, Nano Québec, RMGA, and the Fonds de Recherche Québécois sur la Nature et les Technologies. This work was funded by the Canada Research Chairs program, the Natural Sciences and Engineering Research Council of Canada, the Fonds de Recherche Québécois sur la Nature et les Technologies, and the Fonds de Recherche en Santé du Québec.

CHAPITRE 10

OLIGOMÉRISATION DU SEGMENT POLYGLUTAMINE DE HUNTINGTINE

All-Atom Stability and Oligomerization Simulations of Polyglutamine Nanotubes with and without the 17-Amino-Acid N-terminal Fragment of the Huntingtin Protein

Sébastien Côté†, Guanghong Wei‡, and Normand Mousseau†

† Département de Physique and Groupe d'étude des protéines membranaires (GÉPROM), Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec), Canada

‡ State Key Laboratory of Surface Physics and Department of Physics, Fudan University, 220 Handan Road, Shanghai, 200433, China

Article paru : Côté, S., W. Guanghong, and N. Mousseau. 2012. All-atom stability and oligomerization simulations of polyglutamine nanotubes with and without the 17-amino-acid N-terminal fragment of the huntingtin protein. *J. Chem. Phys. B*. 116 :12168-12179.

10.1 Abstract

Several neurodegenerative diseases are associated with the polyglutamine (polyQ) repeat disorder in which a segment of consecutive glutamines in the native protein is produced with too many glutamines. The Huntington's disease, for example, is related to the misfolding of the Huntingtin protein which occurs when the polyQ segment has more than approximately 36 glutamines. Experimentally, it is known that the polyQ segment alone aggregates into β-rich conformations such as amyloid fibrils. Its aggregation is modulated by the number of glutamine residues as well as by the surrounding amino acid sequences such as the 17-amino-acid N-terminal fragment of Huntingtin which increases the aggregation rate. Little structural information is available, however, regarding the first steps of aggregation and the atomistic mechanisms of oligomerization are yet to be described. Following previous coarse-grained replica-exchange molecular dynamics simulations that show the spontaneous formation of a nanotube consisting of two intertwined antiparallel strands (Laghaei, R. ; Mousseau, N. *J. Chem. Phys.* **2010**, *132*, 165102), we study

this configuration and some extensions of it using all-atom explicit solvent MD simulations. We compare two different lengths for the polyQ segment, 40 and 30 glutamines, and we investigate the impact of the Huntingtin N-terminal residues (htt^{NT}). Our results show that the dimeric nanotubes can provide a building block for the formation of longer nanotubes (hexamers and octamers). These longer nanotubes are characterized by large β-sheet propensities and a small solvent exposure of the main-chain atoms. Moreover, the oligomerization between two nanotubes occurs through the formation of protein/protein H-bonds and can result in an elongation of the water-filled core. Our results also show that the htt^{NT} enhances the structural stability of the β-rich seeds suggesting a new mechanism by which it can increase the aggregation rate of the amyloidogenic polyQ sequence.

10.2 Introduction

Several neurodegenerative diseases are characterized by protein misfolding leading to β-sheet-rich insoluble amyloid deposits in brain tissues [399, 417, 438]. For instance, trinucleotide CAG / polyQ repeat disorders are at the origin of such misfolding for at least nine proteins when the repeat length exceeds a sequence-dependent threshold [26, 168, 358]. Of those nine, the Huntingtin protein misfolds when the native segment of consecutive glutamines at its N-terminal has at least 36 glutamines, forming structures associated with the pathology of the Huntington disease [180]. To unveil the molecular mechanisms behind Huntingtin aggregation, most studies have focused on the Huntingtin exon 1, which can cause neurological phenotype *in vivo* by itself [304], or on fragments of this exon. The Huntingtin exon 1 has an amphipathic N-terminal of 17 residues (htt^{NT}), followed by a segment of consecutive glutamines (polyQ), a segment of consecutive prolines (polyP), and a proline-rich segment.

The aggregation of polyQ alone has been intensively studied [501] as it is the only segment sharing high sequence homology through the nine proteins affected by the polyQ repeat disorder. Early studies postulated different and sometimes conflicting structural models for the fibrils [400]. For instance, X-rays scattering results were interpreted as polyQ aggregating into either β-rich amyloid fibrils characterized by polar zippers [373], or water-filled nanotubes [374]. This latter model was later reinterpreted as stacked β-sheets composed of β-hairpin motifs [432]. Compact β-sheet models characterized by antiparallel strand-turn-strand motifs with each strand being composed of 7 to 9 glutamines were also suggested from muta-

genesis studies [460]. Interestingly, this model induces cell toxicity when enforced in the polyQ sequence of the Huntingtin exon 1 N-terminal [526]. Compatible with the compact β-sheet models, slab-like β-sheet structures were proposed from X-rays scattering experiments [423]. More recently, solid-state NMR spectroscopy suggested that the fibrils of $D_2Q_{15}K_2$ fragments are composed of straight antiparallel β-sheets, while those of $GK_2Q_{38}K_2$ and $GK_2Q_{54}K_2$ peptides are respectively composed of sheet-bend-sheet and sheet-bend-sheet-bend-sheet motifs arranged in superpleated antiparallel cross-β [414].

PolyQ segment as short as 8 glutamines forms mature fibrils *in vitro* [209] through nucleation-elongation process [91], even though the kinetics of fibril formation is repeat-length dependent with longer polyQ segments being more prone to aggregate [90]. The nucleus size is also repeat-length dependent as it is characterized by a sharp transition from nucleus sizes of 4 for Q_{23} to 1 for Q_{26} and longer repeat lengths [222]. It is, however, very challenging to experimentally characterize the structural features of the nucleus and the early on- and off-pathway aggregates to fibrillation. The difficulty resides mainly in the highly dynamical nature of these processes as the oligomeric species exist only in a complex dynamical equilibrium.

To complement experimental observations, these aspects have been investigated by computational studies. For instance, simulations showed that a parallel three-coiled circular water-filled β-helix with 18 or 20 residues per coil is unstable ; while a parallel triangular β-helix with a dry core is stable [446]. Other groups observed the formation and stability of parallel β-helices [97] and β-helical-like monomers [235, 306]. Other simulations showed that water-filled nanotubes made with parallel β-sheets decay into β-helical-like structures characterize by sheet-bend-sheet motifs [520]. Diverse slab-like aggregates of Q_{15} and Q_6 with the polar zipper motif were very stable in all-atom simulations with explicit solvent representation [149]. Nucleation and length-dependent features of polyQ monomers [483], as well as dimerization [485] were also investigated.

The toxicity does not only solely depend on length-dependent structural features of polyQ [245] : the neighboring residues to the polyQ repeat strongly modulate its kinetics of aggregation [349, 396], toxicity [16] and sub-cellular localization [398]. For instance, expressing the polyP sequence of Huntingtin at the C-terminal of polyQ reduces the aggregation propensity [117], while the presence of the first 17 amino acids (htt^{NT}) of Huntingtin at the N-terminal of polyQ enhances the aggregation [275], even in the presence of the C-terminal polyP se-

quence [459]. It has been postulated that the amphipathic character of the htt^{NT} would favor intermolecular interactions between the htt^{NT} bringing the polyQ segments in close contact [457, 459]. Such multi-domain misfolding [409] would enhance the formation of aggregation-prone structural motifs in polyQ. Circular dichroism (CD) spectroscopy studies showed that the htt^{NT} populates α-helical structures [16, 156, 209]. Computational studies on the monomer of htt^{NT} either show very high [233], or moderate [401] α-helical content. When linked to the polyQ segment, the htt^{NT} also populates α-helical structures as shown by X-ray crystallography on exon 1 of Huntingtin [239], and CD experiments on $htt^{NT}Q_N$ aggregates [209]. Computational studies obtained more diverse results : it was shown that the htt^{NT} mostly populates a helix-kink-helix motif in the exon 1 [137], while the $htt^{NT}Q_N$ monomer and dimer adopt rather amorphous configurations when the polyQ fragment consists of more than 20 glutamines [504].

In spite of these efforts, there has been few experimental or computational observations on the structural features of small oligomers of $htt^{NT}Q_N$. In the present study, we investigate such systems – dimers, hexamers and octamers of Q_N and $htt^{NT}Q_N$ for sequences containing 40 and 30 glutamines – using all-atom explicit solvent molecular dynamics (MD) simulations. Starting from the double-stranded antiparallel β-sheet nanotubes obtained previously in unbiased replica exchange molecular dynamics (REMD) simulations [255], we confirm the stability of this fold, and study higher order aggregates composed of the same motif.

This article is constructed as follows. In the next section, we describe our methodology for the various simulations done. We then present our results by starting discussing about the stability of the dimeric nanotube, and the impact of the htt^{NT}. Further, we show that longer nanotubes with and without the htt^{NT} are significantly more stable than their dimeric counterpart, and we describe their growth mechanism. In the last section, we discuss our results in light of previous experimental and computational studies on the polyQ sequence and on the influence of the first 17 residues of the Huntingtin protein.

10.3 Methodology

In this study, we use an all-atom force field with an explicit solvent representation to test the stability of dimeric, hexameric and octameric polyQ nanotubes, and to probe the growth of the hexameric polyQ nanotube. The building block for these folds is taken from unbiased coarse-grained replica exchange molecular dynamics (REMD) simulations performed by our

group [255]. These simulations lead to the spontaneous formation of a dimeric antiparallel polyQ nanotube, which was found to be stable only for chains of more than 35 glutamines. Here, starting from the nanotube obtained for Q_{40}, we investigate its stability at an all-atom level, and compare against similar structures built with Q_{30}. For both chain lengths, we also examine the effect of the Huntingtin protein N-terminal, MATLEKLMKAFESLKSF (residues 1-17, htt^{NT}) on the stability of the nanotube assemblies. These extensive molecular dynamics simulations totalizing 3.7 μs are summarized in Table 10.I.

Table 10.I – Summary of all MD simulations.

Simulation name	Temp. K	Time ns	Box Type	Box Vector nm	No. water
$Q_{40}X2$	300	250	Octahedron	a = b = c = 8	11 566
$htt^{NT}Q_{40}X2$	300	250	Cubic	a = b = c = 12	56 381
$Q_{40}X6$	300	250	Rectangular	a = b = 7, c = 9	13 278
$htt^{NT}Q_{40}X6$	300	250	Cubic	a = b = c = 14	88 662
$Q_{40}X6\_330K$	330	250 X 2	Rectangular	a = b = 7, c = 9	13 278 / 13 286
$Q_{30}X2$	300	250	Octahedron	a = b = c = 7.8	10 762
$htt^{NT}Q_{30}X2$	300	250	Cubic	a = b = c = 12	56 504
$Q_{30}X8$	300	250	Rectangular	a = b = 9, c = 11	28 149
$htt^{NT}Q_{30}X8$	300	250	Cubic	a = b = c = 14	88 409
$Q_{30}X8\_330K$	330	250 X 2	Rectangular	a = b = 9, c = 11	28 149 / 28 131
$Q_{40}X6\_Oligo1$	300	200	Cubic	a = b = c = 11	42 596
$Q_{40}X6\_Oligo2$	300	200	Cubic	a = b = c = 11	42 609
$Q_{40}X6\_Gromos$	330	100	Rectangular	a = b = 7, c = 9	13 329
$Q_{40}X6\_Amber$	330	100	Rectangular	a = b = 7, c = 9	13 312
$Q_{40}X6\_Charmm$	330	100	Rectangular	a = b = 7, c = 9	13 283

We compare two lengths for the polyQ segment : 40 vs. 30 glutamines per polypeptide chain, respectively designated as Q_{40} and Q_{30}. We simulate dimeric nanotubes $Q_{40}X2$ and $Q_{30}X2$, and longer nanotubes – the hexamer $Q_{40}X6$ and the octamer $Q_{30}X8$, both counting 240 glutamines. The impact of the Huntingtin protein N-terminal is studied for every configurations ($htt^{NT}Q_N$). The growth of longer nanotubes is investigated twice in $Q_{40}X6\_Oligo$. Finally, the stability of $Q_{40}X6$ is studied using Gromos53a6, Amber99sb-ILDN or Charmm27 ($Q_{40}X6\_Gromos$, $Q_{40}X6\_Amber$ or $Q_{40}X6\_Charmm$, respectively).

Structures of assemblies. All polyQ folds presented in this study are based upon the results of previous REMD simulations performed with a coarse-grained protein force field [255]. From a completely random initial structure, we observed the spontaneous formation of a nanotube composed of two intertwined Q_{40} polypeptide chains, and characterized by an antiparallel β-sheet with 0.68 H-bond per residue. Reconstruction to all-atom of the coarse-grained side-chains

was performed with SCWRL4 [252], which determines the best side-chain arrangement from the backbone configurations. In the rest of this paper, the reconstructed original nanotube is designated by $Q_{40}X2$. The structure of $Q_{30}X2$ was then obtained from $Q_{40}X2$ by removing 10 glutamine residues at the C-terminal of each chain. Both Q_{40} and Q_{30} dimers were used as building blocks to assemble longer nanotubes : hexamers for Q_{40} and octamers for Q_{30}, both structures consisting of the same number of glutamine residues and winding turns. To assemble these longer nanotubes, the dimers were positioned by hand in such a way as to maximize main-chain/main-chain hydrogen bonding. For the simulations on the impact of the htt^{NT}, we added this 17-residue sequence to the N-terminal of each polypeptide chain in all these structures. The htt^{NT} fragment is initially completely extended without any secondary structure element and it is positioned orthogonally to the nanotube axis.

MD simulations. All simulations are performed in the NPT ensemble using Gromacs version 4.5.4 [42, 185, 278, 478] with the OPLS-AA/L force field [217, 220] and the TIP4P [216] explicit solvent representation. Prior to launch the MD simulations, bad contacts in the reconstructed structures are first removed by energy minimization using the conjugate gradient method with a steepest descent method applied at every 100 steps. Second, structures are solvated and, for the models with the htt^{NT}, two chloride ions per polypeptide chain are added to neutralize the system. Third, the solvated structures are again energy-minimized using the same protocol as described above for the solvent-free case. Fourth, the systems are thermalized at 300 or 330 K for 1 ns using an integration step of 1 fs with position restrains on the protein heavy atoms. Finally, the system is equilibrated at a pressure of 1 atm for 1 ns using an integration time step of 1 fs with position restrains on the heavy atoms of the protein. The conformations at 0 ns in all figures refer to the structures obtained after these pre-MD steps.

Bond lengths of the peptides are constrained with LINCS [184], and water geometries with SETTLE [326], allowing an integration time step of 2 fs. The solute and solvent are separately coupled to external temperature and pressure baths. The temperature of the systems is maintained at 300 or 330 K using the velocity rescaling thermostat [75] with a coupling constant of 0.1 ps. The pressure is isotropically coupled to 1 atm using the Parrinello-Rahman barostat [348, 368] with a coupling constant of 2 ps. The Particle-mesh Ewald (PME) method [116, 150] is used to calculate the electrostatic interactions with a real space cutoff of 1.0 nm. The cutoff is 1.4 nm for the van der Waals interactions. The neighbor list is updated every

20 fs. The center of mass motion of the solute is removed every 10 steps. Conformations are saved every 10 ps. The other relevant parameters such as the temperature, the simulation time, the box type, the box vector lengths, and the number of water molecules for each simulation are shown in Table 10.I.

To ensure that the box size for each simulation is large enough to avoid any self-interaction through the periodic boundary conditions, the minimum distance between the peptide and its adjacent periodic images is always greater than 2 nm. By checking the structural stability of $Q_{40}X6$ using four different force fields : OPLS-AA/L [217, 220], Gromos53a6 [357], CHARMM27 [299, 300], and Amber99sb-ILDN [281], we found that our results are mostly force field independent. As shown in Figure 10.1, the nanotube is stable over 100 ns at the relatively high temperature of 330K for all these force fields. More precisely, the nanotube is most stable when parameterized with the all-atom CHARMM27 force field than with OPLS-AA/L, which we use in the rest of this study, and AMBER99sb-ILDN. The edges of the nanotube are least stable with the united-atom GROMOS53a6 force field. Another simulation on the dimeric nanotube Q40X2 using CHARMM27 also confirms its stability at 300K for 250 ns during which the average β-sheet content is $83\pm3\%$ (vs. $57\pm4\%$ when using OPLS-AA/L).

Analysis. Data analysis is performed with the Gromacs facilities and our in-house codes. To compare the results of Q_{40} with Q_{30}, and to evaluate the impact of the htt^{NT}, we compute a number of quantities : the number of hydrogen bonds, the root mean square deviations (RMSD) with respect to the initial structure, the solvent accessible surface area (SASA) [146], the contact propensities, and the content of secondary structure (using STRIDE [164]). A contact between two non-consecutive residues is considered present when two aliphatic carbons are within 4.8 Å or when any two other atoms are within 5.4 Å of each other [195]. To remove the high frequency thermal fluctuations, all the data reported as a function of time are computed as a running-time average over 5-ns windows.

Convergence is assessed by tracking the backbone root mean square deviation (BB-RMSD) as well as by following other relevant parameters such as the number of hydrogen bonds, the SASA of the backbone atoms, and the secondary structures. We run each simulation long enough to ensure that the last 100 ns is converged as shown in Figure III.1. Accordingly, averages and standard deviations on the BB-RMSD, the percentage of secondary structure, the number of hydrogen bonds, the number of contacts, and the SASA are computed, in all cases, over the in-

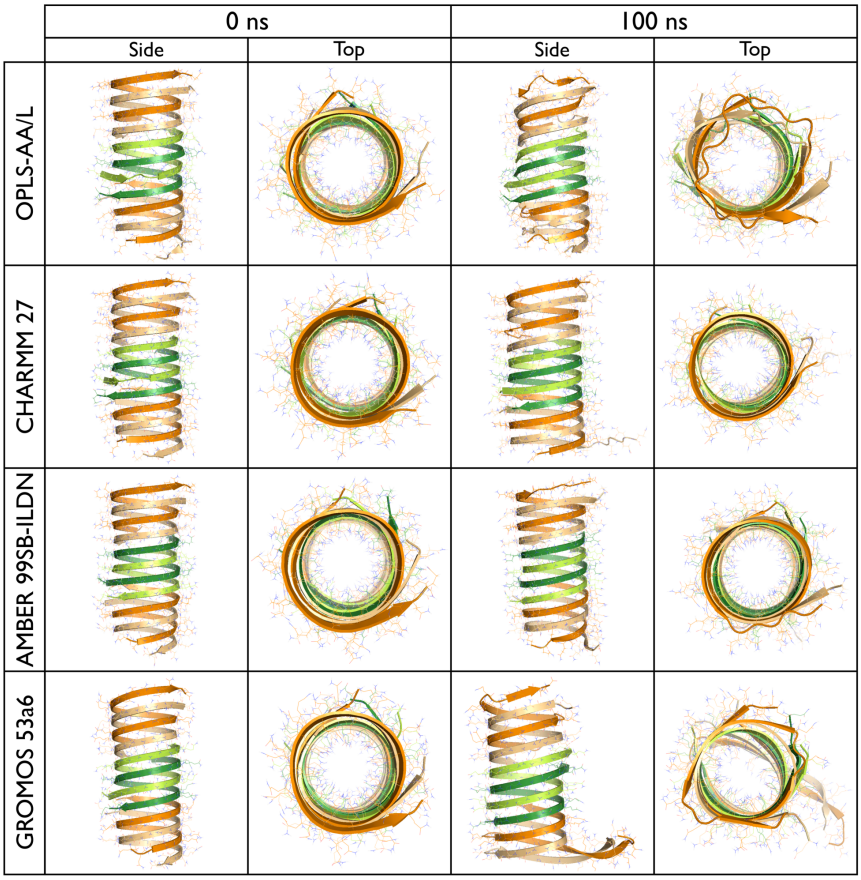


Figure 10.1 – Snapshots of the $Q_{40}X6$ nanotube. The side and top views at 0 and 100 ns using four different force fields at 330 K. From top to bottom, OPLS-AA/L, CHARMM27, AMBER99sb-ILDN and GROMOS53a6. The side and top views are respectively displayed on the left and right columns for each time.

terval 150 to 250 ns. Error bars are the standard deviations of these quantities. Finally, molecular graphics images are generated using the PyMOL software (http ://www.pymol.org).

10.4 Results

10.4.1 Simulations on the dimeric nanotubes

$Q_{40}X2$ and $Q_{30}X2$. The top and side views of the $Q_{40}X2$ and $Q_{30}X2$ nanotubes at 0 and 250 ns are shown in Figure 10.2. After 250 ns, they are still aggregated with no segment interacting preferentially with water molecules. They remain in their original antiparallel arrangement with a significant propensity for β-sheet : $57\pm4\%$ for Q_{40} and $55\pm7\%$ for Q_{30} (Figure 10.3, panel A). As a result, the backbone (BB-) RMSD, computed with respect to the initial structure, converges

at 4.4±0.2 Å for $Q_{40}X2$ and 4.0±0.4 Å for $Q_{30}X2$ (Figure 10.4). The water-filled core of $Q_{40}X2$ collapses at 190 ns due to the formation of protein/protein H-bonds there, while it remains preserved in $Q_{30}X2$. As discussed in the next section, the water-filled core stability increases significantly with the nanotube length.

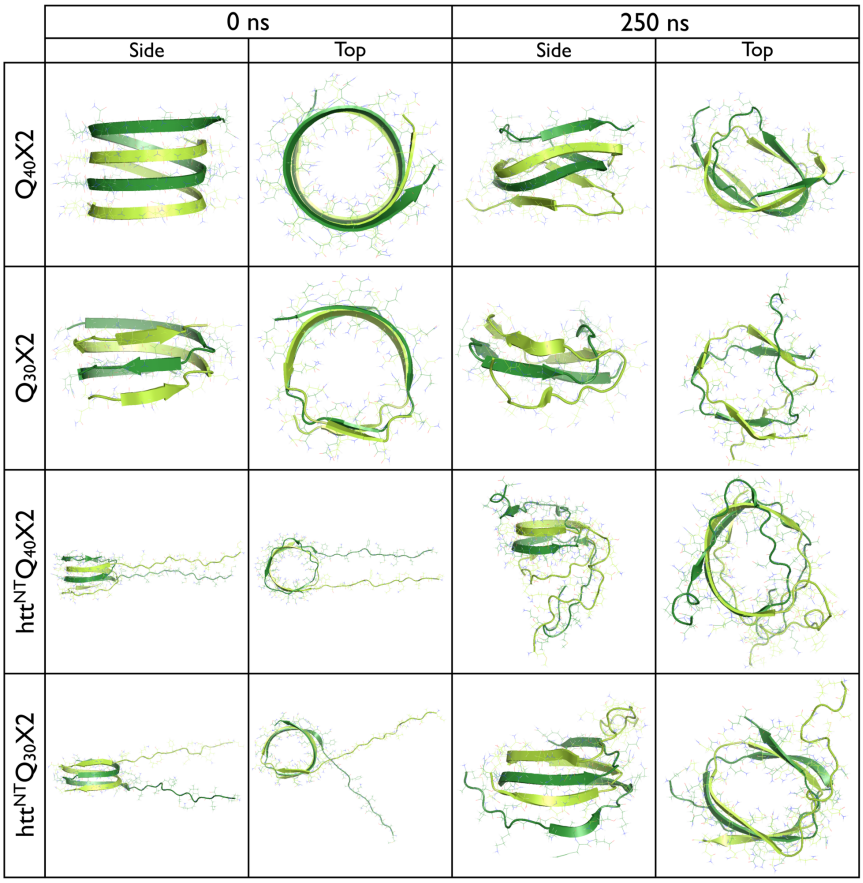


Figure 10.2 – Snapshots of the dimeric nanotubes. The side and top views at 0 and 250 ns of $Q_{40}X2$ and $Q_{30}X2$ are respectively shown on the first and second rows. Structures with the htt^{NT} are shown on the third and fourth rows for $htt^{NT}Q_{40}X2$ and $htt^{NT}Q_{30}X2$, respectively. The side and top views are respectively displayed on the left and right columns for the initial and final states.

To assess the impact of the chain-length on the dimeric nanotubes, we computed quantities related to their stability such as the average β-sheet content, the number of main-chain/main-chain and main-chain/water H-bonds, and the SASA of the main-chain atoms (Figure 10.3). Analysis of these quantities shows only slight differences between $Q_{40}X2$ and $Q_{30}X2$. In terms of the β-sheet content, for example, while the propensity is slightly larger for $Q_{40}X2$, the difference is not significant (57±4% vs. 55±7%, panel A). A similar observation is made on the

average number of main-chain/main-chain H-bonds per residue : it is 0.38±0.03 for Q_{40} compared to 0.35±0.03 for Q_{30} (panel B). In line with these results, the SASA of the main-chain atoms and the number of main-chain/water H-bonds is smaller for Q_{40} (0.156±0.004 nm^2 vs. 0.171±0.003 nm^2 from panel C, and 1.07±0.07 vs. 1.21±0.08 from panel D, respectively). All these data suggest that Q_{40} is slightly more stable than Q_{30}, although this difference is not really significant for the dimers.

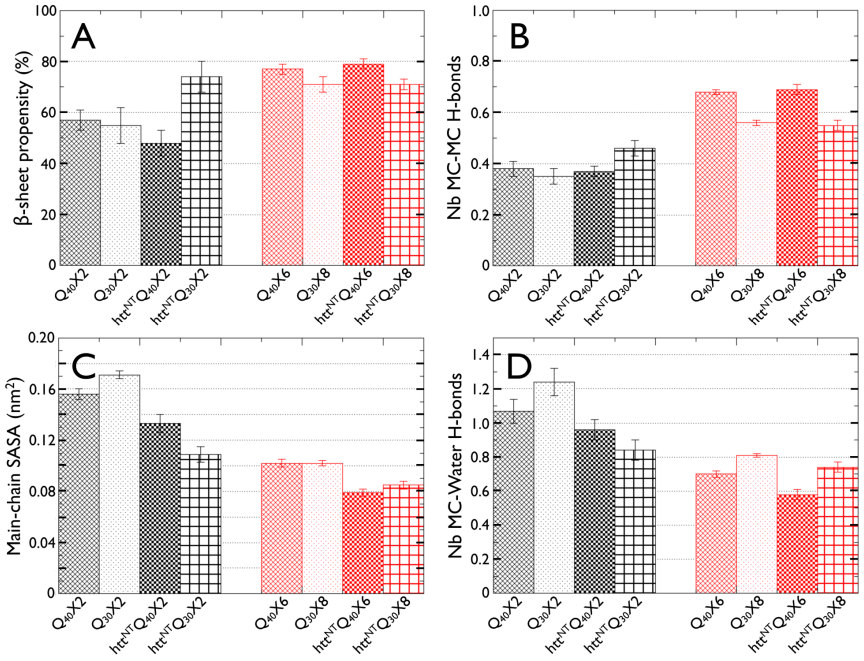


Figure 10.3 – Mean values and standard deviations over the glutamine residues of various structural properties of the nanotubes in the converged time interval (150 to 250 ns). β-sheet propensity (panel A), number of main-chain/main-chain H-bonds per residue (panel B), SASA per residue of the main-chain atoms (panel C), and number of main-chain/water H-bonds per residue (panel D) for the various systems studied here. Higher stability of the nanotube core is correlated to a greater β-sheet content and a larger number of main-chain/main-chain H-bonds, as well as a smaller SASA of the main-chain atoms and a smaller number of main-chain/water H-bonds. The values for the dimers are shown to the left (black), and the values for the longer nanotubes are shown to the right (red). The standard deviations are displayed by error bars.

htt^{NT}Q$_{40}$X2 and htt^{NT}Q$_{30}$X2. We now examine the effect of adding the Huntingtin N-terminal residues, htt^{NT}, to the polyQ chains in Q_{40}X2 and Q_{30}X2. The 250-ns MD simulations show that the addition of the htt^{NT} has a positive impact on the stability of the nanotubes. As for the pure polyQ chains, no part of the polyQ sequence interacts preferentially with water molecules (Figure 10.2). Also, the nanotube core retains the antiparallel β-strand arrangement,

with a significant β-sheet propensity of 48±5% for Q_{40} and 74±6% for Q_{30} (Figure 10.3, panel A). Both the water-filled cores of $htt^{NT}Q_{40}X2$ and of $htt^{NT}Q_{30}X2$ are preserved throughout the simulations, as opposed to $Q_{40}X2$ without the htt^{NT}. The BB-RMSD, calculated on the polyQ region only, is also reduced by 0.5 Å for Q_{40} and by 1.3 Å for Q_{30} (Figure 10.4).

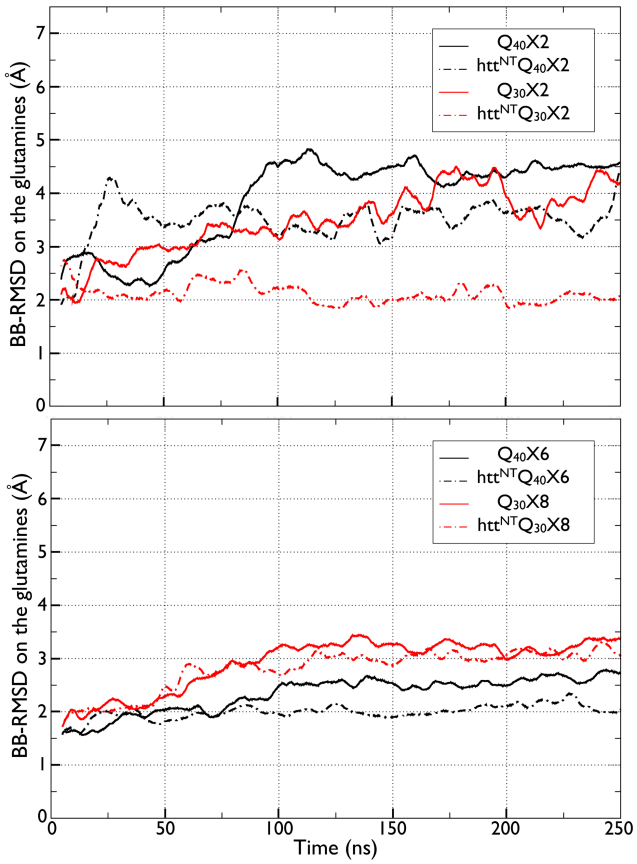


Figure 10.4 – Time evolution of the backbone root mean square deviation (BB-RMSD) on the glutamine residues as measured with respect to the initial structure. The BB-RMSD is computed on the backbone atoms : O, N, C_α and C. The values for the dimers and the longer nanotubes are respectively shown on the top and bottom. The values for Q_{40} and Q_{30} are respectively displayed in black and red. The dotted lines represent the simulations with the htt^{NT}. Data are computed as a running-time average over 5-ns windows.

The stabilizing effect of the htt^{NT} is also seen in other structural quantities. For instance, for the nanotube core, the SASA of the main-chain atoms and the number main-chain/water H-bonds are both reduced when the htt^{NT} is present (panels C and D in Figure 10.3, respectively). In terms of β-sheet content and the number of main-chain/main-chain H-bonds, the Q_{40} nanotube shows little changes after the addition of the htt^{NT}, while these quantities show a significant

increase for Q_{30} (panels A and B in Figure 10.3, respectively). This effect is directly associated with the refolding of the htt^{NT} at the end of the nanotube core, where it forms a β-sheet leading to an extension of the nanotube and fixing the polyQ chains into position. This motif appears very early in the simulation, at 14 ns, and it remains until the end of the run (Figure 10.2, dark green strand at the bottom of $htt^{NT}Q_{30}X2$'s side view at 250 ns).

The htt^{NT} segment itself does not display a well-defined structure. In our simulations, it populates isolated β-bridges, 3-10 helices, turns and random coils, but no α-helix, for both Q_{40} and Q_{30} (Figure III.2). As discussed previously, the htt^{NT} in Q_{30} populates a β-sheet, but not in Q_{40}. In terms of contact propensity, the htt^{NT} of Q_{40} interacts mainly with itself and with the other htt^{NT} (77$\pm$4), and less dominantly with the nanotube core (55$\pm$4). For Q_{30}, the situation is reversed as one htt^{NT} participates in a β-sheet with the nanotube core : there are 44$\pm$3 contacts between themselves, and 92$\pm$4 with the glutamine residues. The htt^{NT} also forms H-bonds with itself and the other htt^{NT} (14$\pm$2 for Q_{40} and 6$\pm$2 for Q_{30}), and with the glutamine residues (7$\pm$2 for Q_{40} and 7$\pm$1 for Q_{30}).

10.4.2 Simulations on the longer nanotubes

$Q_{40}X6$ and $Q_{30}X8$. We also simulated, at 300 and 330 K, longer nanotubes assembled from the dimers : hexamers for Q_{40} and octamers for Q_{30}, both having 240 glutamines and the same number of winding turns. These structures are described in the Methodology section.

First, we present the simulations at 300 K and compare them to the dimers. After 250 ns, we clearly see that the longer nanotubes $Q_{40}X6$ and $Q_{30}X8$, with their water-filled core remaining intact throughout the simulations, are significantly more stable than their dimeric counterpart (Figure 10.5 vs. Figure 10.2). Clearly, the BB-RMSD, computed with respect to the initial structure, supports this observation as it converges at 2.6$\pm$0.1 Å for Q_{40} and 3.2$\pm$0.2 Å for Q_{30}, which are smaller than the dimers' RMSD values by about 1-2 Å (Figure 10.4, bottom vs. top panel). The propensity of β-sheet is also significantly larger for the longer nanotubes with 80$\pm$2% for Q_{40} and 73$\pm$3% for Q_{30} (Figure 10.3, panel A). Moreover, the main-chain/main-chain H-bonds network, which is characteristic of the nanotube stability, is more protected from solvent when compared to the dimers. This is shown by a larger number of main-chain/main-chain H-bonds (panel B), a smaller SASA of the main-chain atoms (panel C), and a smaller number of main-chain/water H-bonds (panel D).

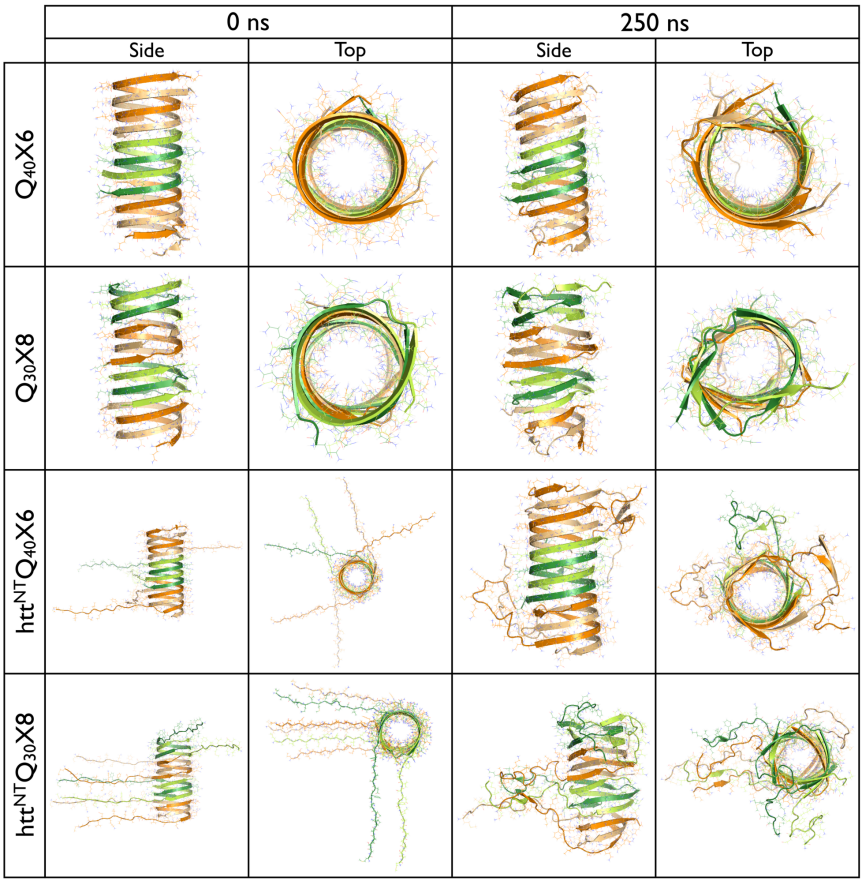


Figure 10.5 – Snapshots of the longer nanotubes. The side and top views at 0 and 250 ns of $Q_{40}X6$ and $Q_{30}X8$ are respectively shown on the first and second rows. Structures with the htt^{NT} are shown on the third and fourth rows for $htt^{NT}Q_{40}X6$ and $htt^{NT}Q_{30}X8$, respectively. The side and top views are respectively displayed on the left and right columns for the initial and final states.

Comparing Q_{40} and Q_{30} more closely, we observe significant differences in terms of structural stability. As shown previously, the BB-RMSD is smaller for Q_{40} than Q_{30} (2.6±0.1 Å vs. 3.2±0.2 Å, bottom panel in Figure 10.4), and Q_{40} has a larger β-sheet propensity than Q_{30} (80±2% vs. 73±3%, Figure 10.3 on panel A). Also, the main-chain of Q_{40} forms more H-bonds with the other main-chains than Q_{30} (0.68±0.01 vs. 0.56±0.01 from panel B), and less H-bonds with water molecules (0.70±0.02 vs. 0.81±0.01 from panel D). As a result, the nanotube structure $Q_{40}X6$ is more ordered than that of $Q_{30}X8$ (Figure 10.5).

Further analysis at 330 K confirms the stability of the longer nanotubes. We did two independent simulations of 250 ns at 330 K for both $Q_{40}X6$ and $Q_{30}X8$. For each nanotube, two different initial structures were used : the same nanotubes as for the simulations at 300K, and a

new nanotube for each chain-length, assembled according to the prescription detailed in the Methodology section. The all-atom RMSD between the two initial structures is 3.8 Å for Q_{40} and 0.9 Å for Q_{30}, which is due to each dimer being rotated slightly differently with respect to each other along the nanotube central axis. These simulations confirm that, even at a relatively high temperature, the longer nanotubes are stable (Figure 10.6) as shown by the small BB-RMSD, which stays below 5 Å (Figure III.3). While the BB-RMSD are comparable for Q_{40} and Q_{30}, the latter structure is considerably more distorted and its water filled core is almost collapsed at 250 ns. For Q_{40}, some residues at one edge are detached from the nanotube after 200 ns due to the larger thermal fluctuations at the termini, while the middle subunit is very stable and undergoes small thermal fluctuations.

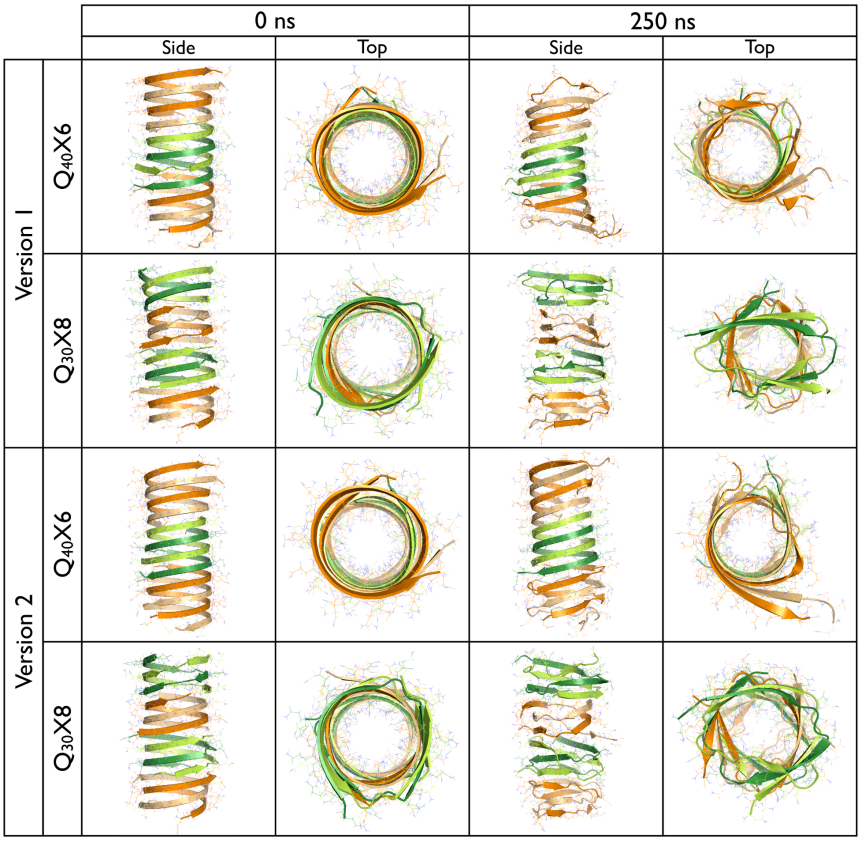


Figure 10.6 – Snapshots of the longer nanotubes at 330 K. The side and top views at 0 and 250 ns of the simulations on $Q_{40}X6$ and $Q_{30}X8$ starting from two different initial structures, versions 1 and 2. The side and top views are respectively displayed on the left and right columns for the initial and final states.

htt^{NT}Q$_{40}$X6 and htt^{NT}Q$_{30}$X8. After 250 ns, the longer nanotubes of Q_{40} and Q_{30} with the added htt^{NT} are still aggregated (Figure 10.5). Here again, the addition of the htt^{NT} plays

a stabilizing role for the nanotube : the water-filled core remains intact, deformation over the simulation time-scale is minimal, and the BB-RMSD, computed from the initial structure on the glutamine residues, remains low (Figure 10.4, bottom panel). For Q_{40}, the BB-RMSD plateaus at 2.0±0.1 Å, compared to 2.6±0.1 Å for the pure polyQ sequence, while it remains mostly unchanged for Q_{30} (3.2±0.2 Å compared to 3.1±0.1 Å). The main-chain is also less accessible to the solvent (Figure 10.3, panel C), forming fewer hydrogen bonds with the water molecules (panel D). Internally, however, the structure is not affected by the addition of the htt^{NT} : the β-sheet propensity (panel A) and the number of main-chain/main-chain H-bonds (panel B) are essentially identical with and without it. Overall, the htt^{NT} stabilizes the longer nanotubes core by reducing the solvent exposure of the glutamines and the thermal fluctuations of the termini.

The structure of the htt^{NT} in the presence of the longer nanotubes is similar to that observed in the dimers. The segment populates β-sheet, isolated β-bridge, turn and random coil for both $htt^{NT}Q_{40}X6$ and $htt^{NT}Q_{30}X8$ (Figure III.4). A transient short 3-10 helix is also seen for Q_{40}, but no α-helix is observed over the time scale of our simulations. In terms of contact, the htt^{NT} interacts mostly intra- and intermolecularly with itself (202±7 for Q_{40} and 325±10 for Q_{30}). Contacts are also observed, but to a lesser extend, between the htt^{NT} and the glutamine residues forming the nanotube core (200±8 for Q_{40} and 205±9 for Q_{30}). On average, the htt^{NT} forms H-bonds in equal number with itself and with the others htt^{NT}, and with the glutamines for Q_{40} (26±4 and 26±3, respectively), while it much prefers self-contacts for the shorter polyQ sequence (40±4 vs. 25±3), probably due to the higher ratio of htt^{NT} to the glutamines in the latter.

10.4.3 Formation of the longer nanotube

In the previous sections, we observed that the nanotube stability is significantly increased for longer nanotubes. We now investigate the process by which the nanotube $Q_{40}X6$ could be assembled from dimeric nanotubes in solution. We perform two independent simulations of 200 ns at 300 K on a system composed of a dimeric and a tetrameric nanotube, the latter providing a weakly fluctuating seed to study oligomerization. These two simulations are launched from two different initial structures for which the position and the orientation of the dimer is varied with respect to the tetramer.

In the first system, the two nanotubes are positioned such as to avoid direct interactions

between the dimer and the tetramer in the initial state, $Q_{40}X6$_Oligo1. The minimum distance between the two aggregates is 1.8 nm and their respective nanotube axis is perpendicular to one another as shown in Figure 10.7. In the same figure, we observe that at 63 ns the dimer binds to the tetramer through the formation of a single H-bond between the NH_2 group of a glutamine side-chain in the dimer and the C=O group of a glutamine main-chain in the tetramer. Once this H-bond forms, others are rapidly created between the dimer and the tetramer (Figure 10.8) aligning the axis of the dimeric nanotube toward that of the tetramer. The number of H-bonds between the two nanotubes increases to 7 at 100 ns. After 200 ns, the dimer and the tetramer share 18 H-bonds, the angle between their axis is $\sim 50^{\circ}$, and the total β-sheet propensity is 70%. We expect that longer simulations would allow further reorganizations to occur toward the nanotube $Q_{40}X6$, a very stable structure as discussed in the previous section (Figure 10.5) with a β-sheet propensity of $80 \pm 2\%$ (Figure 10.3, panel A).

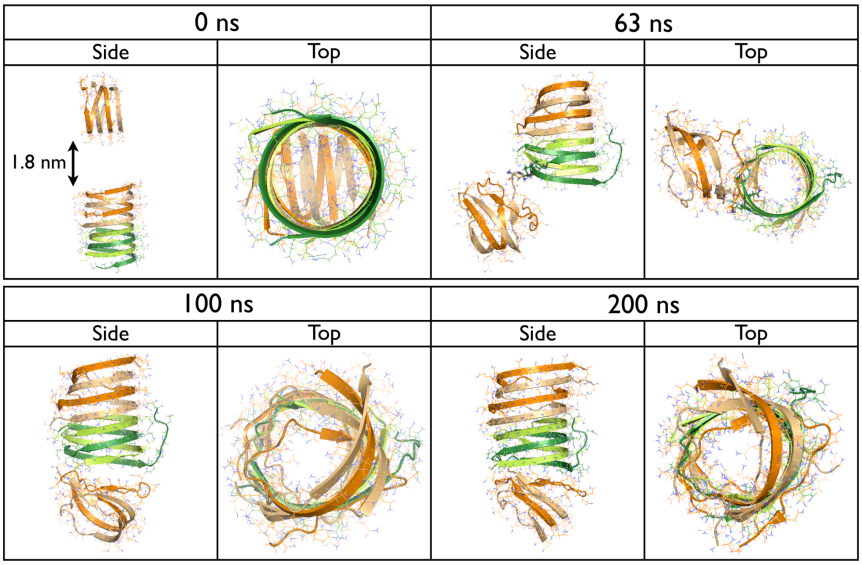


Figure 10.7 – Snapshots during the $Q_{40}X6$_Oligo1 simulation. The side and top views at 0, 63, 100 and 200 ns during the oligomerization between a dimeric and a tetrameric nanotube. Initially, the two nanotubes are positioned at a minimum distance of 1.8 nm. At 63 ns, the dimer binds to the tetramer through the formation of a single H-bond, and further alignment of the dimeric nanotube axis to that of the tetramer follows.

For the second simulation, $Q_{40}X6$_Oligo2, we also ensure that there is no interaction between the dimeric and tetrameric nanotube initially, with a minimum distance of 1.8 nm between the two nanotubes. This time, however, they are positioned along the same axis and placed on top of each other (Figure III.5). At 40 ns, the dimer binds to the side of the tetramer nanotube

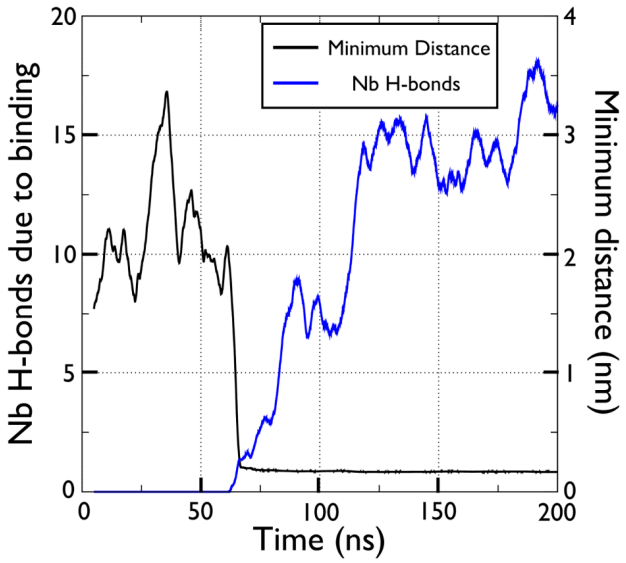


Figure 10.8 – Time evolution of the number of hydrogen bonds between the dimeric and the tetrameric nanotube during the $Q_{40}X6$_Oligo1 simulation. The number of hydrogen bonds (left axis, blue) is correlated to the minimal distance between the dimer and the tetramer (right axis, black).

through the formation of side-chain/side-chain H-bonds (Figure III.6). At 200 ns, the dimer is still attached on the side of the tetramer and both structures share 12 H-bonds, which is 6 H-bonds less than in $Q_{40}X6$_Oligo1. The total β-sheet propensity, however, is similar in both simulations (71% vs. 70%). Unfortunately, it is not possible to quantitatively determine if one of the binding modes – on the top or on the side of the nanotube – is energetically more favorable from our standard molecular dynamics simulations.

10.5 Discussion

10.5.1 The nanotube configuration

In a previous study, we observed the spontaneous formation of a nanotube made of two anti-parallel intertwined strands using replica exchange molecular dynamics (REMD) simulations coupled to a coarse-grained force field (Q_{40} at 0 ns in Figure 10.2) [255]. This structure is intrinsically different from the water-filled nanotube structure previously considered that is made of parallel non-intertwined strands [374]. Molecular dynamics simulations showed that the parallel nanotube model is unstable for the monomer [320], and can decay toward β-helical-like structures with sheet-bend-sheet motifs [520]. Even if oligomeric parallel nanotubes are more

stable, they still undergo large distortions within short time scale molecular dynamics (MD) simulations [336, 352]. Here, we show that the nanotubes made of antiparallel intertwined strands are stable over long time-scales using all-atom MD simulations with explicit solvent.

Indeed, the dimers remain folded for 250 ns (Figure 10.2) as shown by the small backbone root mean square deviation (BB-RMSD) with respect to the initial structure (Figure 10.4, top panel), and the large β-sheet propensity above 50% (Figure 10.3, panel A). While the dimers undergo some distortions, the longer nanotubes are significantly more stable (Figure 10.5) as shown by reduced BB-RMSD values (Figure 10.4, bottom panel), even at 330 K (Figure 10.6) and independently of the force field used (Figure 10.1). These longer nanotubes have also a larger β-sheet propensity ($80\pm2\%$ for $Q_{40}X6$, and $73\pm3\%$ for $Q_{30}X8$ from Figure 10.3 on panel A).

Recently, molecular structures of the $D_2Q_{15}K_2$, $GK_2Q_{38}K_2$ and $GK_2Q_{54}K_2$ fibrils have been proposed based on solid-state NMR spectroscopy data [414]. It was suggested that each strand is either extended ($D_2Q_{15}K_2$), extended-bend-extended ($GK_2Q_{38}K_2$), or extended-bend-extended-bend-extended ($GK_2Q_{54}K_2$). These motifs are arranged in antiparallel fashion along the fibril axis, an arrangement that shares structural features with our nanotube model. First, both structures are composed of antiparallel β-sheets. Second, our nanotube is made of sheet-bend-sheet motifs with each sheet consisting of $\sim$17 residues (Figure III.4), which is similar to the $\sim$15 residues observed by NMR and to the 7-9 residues hypothesized from mutagenesis results [460]. Finally, the intertwined nanotube is similar to an out-of-registry fibril with a sheet-bend-sheet motif.

Whereas the fibril structures determined by solid-state NMR spectroscopy [414] and X-rays scattering experiments [373, 423] are characterized by steric zipper structures with a dry core, the nanotubes studied here possess a water-filled core. In spite of these differences, these structures are stable at both 300 K and 330 K for 250 ns (Figures 10.5 and 10.6, respectively). Moreover, the small variations shown by different structural properties of the nanotubes, particularly for the longer ones, over the last 100 ns confirm the convergence of our simulations (Figure III.1). Other all-atom MD simulations with explicit solvent show that the steric zipper structure with a dry core is also stable, at least over 50 ns [149]. Taken together, these results suggest that there exists a possible structural diversity for polyQ aggregates, at least for oligomers. Such structural diversity was previously confirmed for other amyloid sequences, which

are expected to share high structural similarities despite their low sequence homology [417]. For instance, the amyloid-beta (Aβ) protein was shown to populate a wide range of fibrillar structures [367, 376, 469]. Interestingly, a water-filled core structure was recently observed for Aβ_{1-42} using Cryo-electron microscopy (CryoEM) [527] and simulations [321]. As amyloid proteins are expected to share common structural features, water-filled structures cannot be ruled out for polyQ aggregates, and our results show that at least one nanotube structure – intertwined antiparallel strands – is stable. Of course, the polymorphism of Aβ has been associated essentially with different quaternary arrangement of monomers with the same sheet-turn-sheet motif with a steric zipper [321, 366]. This secondary structure appears driven by the many hydrophobic residues that need to be buried from the solvent. For polyQ, the polymorphism could be more diverse as it is exclusively composed of hydrophilic glutamines which can form H-bonds either with other glutamines or with water molecules.

The growth of the longer nanotubes occurs via the formation of a single H-bond between two shorter nanotubes followed by the formation of other H-bonds favouring the alignment of the nanotubes along the tube axis (Figures 10.7 and 10.8). The shortest subunit of these longer assemblies is a dimeric nanotube whose formation was previously shown to occur in four main steps [255]. (1) Apparition of extended regions in each monomer ; (2) formation of antiparallel β-sheets ; (3) structural reorganization occurs toward a water-filled triangular structure with sheet-bend-sheet motifs ; (4) rearrangement into a nanotube as the triangular structure becomes cylindrical. Formation of extended regions in the monomers during the first step is critical to the formation of the nanotube. Interestingly, other simulations [483] also show that the formation of extended structures in the Q_{40} monomer is energetically unfavorable, agreeing with the monomeric nucleus size of 1 for Q_{40} observed experimentally [222]. It is also compatible with a recent computational study showing that monomeric parallel nanotubes and steric zippers are not stable and that polyQ favors antiparallel conformations, such as β-sheet stacks [320].

It is also possible, on the other hand, that the nanotube structure is formed independently from the monomeric nucleus for two reasons. (1) It is not clear yet if the formation of polyQ oligomers can preclude nucleation or not [209, 266, 501], and (2) fibril-incompetent oligomers could also be populated during fibrillation, explaining the non-integer and negative nucleus sizes of Q_N and $htt^{NT}Q_N$ previously observed [484].

10.5.2 Effect of the chain length (Q_{40} vs. Q_{30})

All polyQ repeat disorder diseases are characterized by a length-dependent pathological threshold on the polyQ segment [180]. The disease occurs only for the protein containing a certain number of consecutive glutamines, and it occurs faster as the polyQ segment is longer. As the threshold depends on the native protein, being approximately 36 glutamines for Huntingtin, it can not solely originate from the polyQ segment alone [245]. In fact, β-sheet-rich fibrils are observed *in vitro* for polypeptide chains having as few as 8 glutamines [209], with the aggregation being faster for longer polyQ segment [91].

Previously, using a coarse-grained model, we showed that the nanotube composed of two intertwined antiparallel chains, which can spontaneously form for Q_{40}, is unstable when there is less than 35 glutamines [255]. Here, we reinvestigated the stability of the Q_{30} dimeric nanotube using all-atom MD simulations with explicit solvent. In our simulations, Q_{30}X2 does not unfold after 250 ns and keeps its water-filled core (Figure 10.2), showing that water molecules and side-chain H-bonds are very important for stabilizing the nanotube structure of smaller chain-lengths. The dimeric nanotube of Q_{30} also keeps a large number of main-chain/main-chain H-bonds (Figure 10.3, panel B), and a high β-sheet propensity ($55\pm7\%$, from Figure 10.3 on panel A). For Q_{40}X2, these quantities are slightly larger suggesting that the longer sequence forms a more stable nanotube. While this observation is not significant for the dimers, it is for the longer nanotubes Q_{40}X6 and Q_{30}X8 (Figure 10.3). Q_{40}X6 is also more ordered than the shorter sequence at both 300 K (Figure 10.5) and 330 K (Figure 10.6).

Our results are in agreement with other computational studies showing that longer polyQ segments are more ordered and favor dimerization. For instance, replica-exchange discrete molecular dynamics (RE-DMD) simulations coupled to a coarse-grained force field show that the formation of intermolecular H-bonds between polyQ segments is facilitated with longer repeat length in Chymotrypsin inhibitor 2–polyQ chimeras [23]. Another DMD coarse-grained study on 24-mers of polyQ segment of diverse lengths showed that the β-sheet propensity and the aggregate stability increase for longer chain-length [306]. These 24-mers formed annular, ring and tube-like configurations [305, 306]. Finally, Monte-Carlo simulations performed with an all-atom, implicit solvent force field showed that the probability of homodimerization of polyQ segments increases for longer chain-length [485], even though the sampled structures were mostly amorphous with very low secondary structures propensities.

Overall, simulations tend to show that both oligomerization and stability are enhanced with a larger number of glutamines in the polyQ segment is larger in agreement with the increased rate of aggregation observed experimentally for longer repeat-length [91]. These previous conclusions also apply to the intertwined antiparallel nanotube studied here.

10.5.3 Effect of the htt^{NT}

It is known that residues adjacent to the polyQ segment in the native protein strongly modulate the kinetics of aggregation [349]. For instance, the presence of the first 17 amino acids of Huntingtin (htt^{NT}), an amphipathic segment, at the N-terminal of polyQ enhances oligomerization [16, 398] by favouring the appearance of intermolecular contacts between polyQ segments [209, 275].

Here, we investigated the effect of the htt^{NT} on the dimeric nanotubes $Q_{40}X2$ and $Q_{30}X2$, as well as on the longer nanotubes $Q_{40}X6$ and $Q_{30}X8$. The htt^{NT} increases the dimers' stability by reducing the solvent accessibility of the glutamines residues (Figure 10.3 on panels C and D) and the thermal fluctuations at the polyQ N-terminal as shown by the smaller BB-RMSD (Figure 10.4). It also helps to preserve the water-filled core of both $Q_{40}X2$ and $Q_{30}X2$. Interestingly, for the 30-residue polyQ sequence, the htt^{NT} forms a β-sheet with the glutamine residues of the nanotube core at 14 ns that remains present at 250 ns (Figure 10.2, dark green strand at the bottom of htt$^{NT}Q_{30}$'s side view at 250 ns). This motif greatly stabilizes the nanotube core as shown by a significant increase of β-sheet propensity (55$\pm$7% to 74$\pm$6%). Even if a similar motif is not observed for our Q_{40} simulations, nothing prevents it from forming and helping to stabilize the nanotube structure during assembly. Similarly to the dimers, the longer nanotubes are also stabilized by the htt^{NT} (Figures 10.4 and 10.5) through a reduction of the solvent accessibility of the glutamine residues (Figure 10.3 on panels C and D) and the thermal fluctuations at the polyQ N-terminal.

In terms of secondary structures, the htt^{NT} shows a strong structural diversity as it populates turn, β-sheet, isolated β-bridge and random coil in all structures (Figures III.2 and III.4). We do not observe any α-helix on the timescale of our simulations, while CD spectroscopy experiments, for example, show that the htt^{NT} populates about $\sim$45% helical configurations [16]. Aggregates of htt$^{NT}Q_N$ are also found to populate α-helices with htt$^{NT}Q_{35}K_2$ having a percentage of $\sim$50% [209]. Computational results are less conclusive, however. Replica exchange molecu-

lar dynamics (REMD) simulations combined with the AMBER03 all-atom implicit solvent force field show that the htt^{NT} in $htt^{NT}Q_{17}P_{11}$ and $htt^{NT}Q_{55}P_{11}$ mostly populates α-helices (50-80%) at 301 K, as well as 3-10 helices and turns to a lesser extent [137]. Similar results were obtained using simulated tempering (ST) simulations coupled to the AMBER03 force field on the htt^{NT} headpiece alone [483]. These simulations showed that it populates α-helices (20-80%) at 300 K, as well as 3-10 helices and turns to a lesser extent. On the other hand, bias exchange metadynamics (BEM) simulations with the all-atom AMBER99 force field and explicit solvent show that the htt^{NT} is mostly random coil with about $\sim$29% α-helices [401]. A replica-exchange discrete molecular dynamics (RE-DMD) simulations coupled to a coarse-grained model even showed that the htt^{NT} populates more extended (20-30%) than α-helical (5-10%) configurations in the exon 1 of Huntingtin [260]. Mostly disordered aggregates were also observed for $htt^{NT}Q_N$ using Monte Carlo simulations coupled to an all-atom force field with implicit solvent [504]. Overall, most studies tend to show that the htt^{NT} is structurally diverse, and populates α-helices to an extent that varies from one study to the other.

In line with those results, the htt^{NT} in our simulations populates different configurations such as 3-10 helix, turn and random coil. However, it does not populate any α-helical configurations. To assess whether this absence is due to the presence of the nanotubes or is due to the sequence itself, we run a MD simulation at 300 K on a htt^{NT} monomer starting from a single α-helix configuration (data not shown). A running-time average on the percentage of α-helical content shows that the probability drops rapidly and remains low ($< 5\%$) until 60 ns at which point the helix is completely unfolded. The probability of α-helix remains at 0% to the end of the simulation at 150 ns. The β-sheet propensity is also at 0% throughout the simulation. On the contrary, the htt^{NT} remains in a single alpha-helix to the end of the simulation at 150 ns when using CHARMM27 (data not shown), a force field that is known to bias toward alpha-helical configurations [162, 163]. Some of the computational studies mentioned earlier [137, 483] were also done using a force field that is known to bias towards α-helical configurations, AMBER03 [323]. The complete determination of the structure of the htt^{NT} fragment will require a thorough investigation using new force fields such as CHARMM22* and AMBER99sb-ILDN* that have a better secondary structure balance [46, 279, 382].

Whether or not our selected force field underestimates the presence of α-helical structures in the htt^{NT}, our results show that this segment stabilizes the dimeric nanotube model studied here,

which is composed of two intertwined antiparallel strands, and the longer nanotubes composed of these. The htt^{NT} could therefore favor oligomerization not only by bringing the glutamine segments close together as suggested from experiments, but also by enhancing the stability of the β-sheet-rich seed as oligomerization occurs. It would be interesting to verify if such assertion is also observed for the polar zipper structure having the sheet-bend-sheet motif as recently determined by solid-state NMR spectroscopy [414].

10.6 Conclusion

Although the existence of water-filled polyQ nanotubes is certainly not fully accepted today, growing evidence of polymorphism for other amyloid proteins such as amyloid-beta however suggests that this structure cannot yet be ruled out [321, 376, 527]. Unbiased simulations have shown the spontaneous formation of a double-stranded antiparallel β-sheet nanotube [255]. The all-atom MD simulations with explicit solvent presented here show that this structural model serve as a building block to form longer nanotubes with higher structural stability. Molecular dynamics simulations for stability check using various well-known force fields confirm that these results are not an artifact. Moreover, all assemblies are stabilized by the first 17 amino acids of the Huntingtin protein (htt^{NT}), enhancing the stability of the β-sheet-rich seeds as oligomerization occurs. This latter observation suggests that the role of the htt^{NT} segment in increasing the aggregation rate of polyQ may be more diverse than previously thought, and shows that more experimental and computational studies are needed to better characterize its role at the atomic level.

10.7 Supporting Material

In Annexe III, figures showing the time evolution of various structural properties of the nanotubes, per residue secondary structures of the dimeric nanotubes, time evolution of the BB-RMSD of the longer nanotubes at 330K, per residue secondary structures of the longer nanotubes, snapshots during the Q_{40}X6_Oligo2 simulation and time evolution of the number of H-bonds during the Q_{40}X6_Oligo2 simulation.

10.8 Author contributions

S.C., G.W. and N.M. designed the research ; S.C. performed the research ; S.C. contributed analytic tools ; S.C., G.W. and N.M. analyzed the data ; and S.C., G.W. and N.M. wrote the manuscript.

10.9 Acknowledgments

This work was funded in part by the Canada Research Chairs program, the Fonds québécois de recherche sur la nature et les technologies (FQRNT), the Natural Sciences and Engineering Reasearch Council of Canada (NSERC) and the Fonds de recherche en santé du Québec (FRSQ). G. W. acknowledges the financial support from the National Natural Science Foundation of China (Grant No. 11074047). Calculations were done on the supercomputers of Calcul Québec. The authors thank Rozita Laghaei for helpful discussions.

CHAPITRE 11

INTERACTIONS MEMBRANAIRES DU N-TERMINAL DE HUNTINGTINE – PARTIE 1

Atomistic mechanisms of huntingtin N-terminal fragment insertion on a phospholipid bilayer revealed by molecular dynamics simulations

Sébastien Côté†, Guanghong Wei‡, and Normand Mousseau†

† Département de Physique and Groupe d'étude des protéines membranaires (GÉPROM), Université de Montréal, Montréal (Québec), Canada

‡ State Key Laboratory of Surface Physics and Department of Physics, Fudan University, Shanghai, 200433, China

Article paru : Côté, S., W. Guanghong, and N. Mousseau. 2014. Atomistic mechanisms of huntingtin N-terminal fragment insertion on a phospholipid bilayer revealed by molecular dynamics simulations. *Proteins*. 82 :1409-1427.

11.1 Abstract

The huntingtin protein is characterized by a segment of consecutive glutamines (Q_N) that is responsible for its fibrillation. As with other amyloid proteins, misfolding of huntingtin is related to Huntington's disease through pathways that can involve interactions with phospholipid membranes. Experimental results suggest that the N-terminal 17-amino-acid sequence (htt^{NT}) positioned just before the Q_N region is important for the binding of huntingtin to membranes. Through all-atom explicit solvent molecular dynamics simulations, we unveil the structure and dynamics of the $htt^{NT}Q_N$ fragment on a phospholipid membrane at the atomic level. We observe that the insertion dynamics of this peptide can be described by four main steps – approach, reorganization, anchoring and insertion – that are very diverse at the atomic level. On the membrane, the htt^{NT} peptide forms a stable α-helix essentially parallel to the membrane with its non-polar side-chains – mainly Leu-4, Leu-7, Phe-11 and Leu-14 – positioned in the hydrophobic core of the membrane. Salt-bridges involving Glu-5, Glu-12, Lys-6 and Lys-15, as well as hydrogen bonds involving Thr-3 and Ser-13 with the phospholipids also stabilize the structure and orientation of the htt^{NT} peptide. These observations do not significantly change upon adding the Q_N

region whose role is rather to provide, through its hydrogen bonds with the phospholipids' head group, a stable scaffold facilitating the partitioning of the htt^{NT} region in the membrane. Moreover, by staying accessible to the solvent, the amyloidogenic Q_N region could also play a key role for the oligomerization of $htt^{NT}Q_N$ on phospholipid membranes.

11.2 Introduction

The huntingtin protein belongs to a family of 10 proteins that are related to trinucleotide CAG / polyglutamine repeats disorder [26, 168, 358]. These proteins are characterized by a consecutive segment of glutamines of varying lengths that can lead, when the number of repeats exceeds a given threshold, to misfolding and fibrillation. Misfolding of huntingtin is related to the Huntington's disease and it occurs when its consecutive segment of glutamines reaches 35, with the neurotoxicity strength likely being correlated to this length [180]. The origin of the neurotoxicity is still under debate, but experimental evidences suggest that it could involve non-physiological interactions of huntingtin with membrane-containing organelles of the cell [16, 186, 219, 327, 398], a characteristic shared by many other neurodegenerative diseases [263, 264]. Most experiments and simulations focus on huntingtin exon 1 or some of its fragments since these are sufficient to induce a pathology *in vivo* similar to the full-length protein [304, 350, 526]. Huntingtin exon 1 consists of a 17-residues amphipathic N-terminal (htt^{NT}), a consecutive segment of glutamines (Q_N), a consecutive segment of prolines (P_N) and a proline-rich segment.

As other amyloid proteins, huntingtin can aggregate into fibrils through a nucleation-dependent process [400, 414, 423, 501]. Its kinetics of aggregation is dependent on the number of glutamines [91] as well as on the presence of the neighboring segments [349, 396] such as htt^{NT} and P_N. For instance, the nucleus size of Q_N is reduced from 4 to 1 when the number of glutamines passes from 23 to 26 [222], while the presence of the htt^{NT} segment leads to a negative nucleus size suggesting spontaneous fibrillation [459] or limitations of the homogenous nucleation model [484]. The aggregation kinetics is also sequence-dependent since fibrillation is faster when the number of glutamines in the Q_N region increases [90, 91] and when the htt^{NT} region is present [275, 459], while the fibrillation kinetics is slowed down by the presence of the P_N region [51, 117].

Since it is difficult to resolve directly the atomic structure of huntingtin's transient monomer

and small oligomers in solution due to fast aggregation, most structural information comes from solution circular dichroism (CD) experiments. These measures show that the htt^{NT} peptide in solution populates a distribution of α-helical content that varies from 10 to 55% [16, 319, 459, 504]. Nuclear magnetic resonance measurements on the structural ensemble of the htt^{NT} peptide suggest, for their part, that it does not form stable α-helices [459]. X-rays scattering on a fusion protein of maltose-binding protein and the exon 1 of huntingtin (MBP-$htt^{NT}Q_{17}$-ex1) shows that the htt^{NT} segment folds into an α-helix, while the Q_N segment appears to be unstructured [239] as also shown by fluorescence correlation spectroscopy [113].

Computational studies have partially complemented these experiments by investigating the structures and dynamics of monomer/oligomers folding at the atomic level. For instance, simulations on the monomeric and dimeric Q_N peptides of various lengths revealed that they form mostly disordered, collapsed globules suggesting that the formation of β-sheets is more favorable in aggregates of higher molecular weights [483, 485]. The Q_N peptide can nevertheless also adopt various stable motifs at the monomer level such as α-helix, β-sheet and β-sheetstack [320] as well as β-helix [97, 446]. Moreover, the β-sheet stability increases as Q_N is longer [235, 255, 260, 306] and as the oligomeric weight increases [110, 306, 336]. Other simulations were performed to discriminate the widely different repeat motifs suggested for the fibril morphologies of Q_N [149, 352, 520]. The effect of neighboring regions on Q_N folding has also been investigated. For instance, simulations on htt^{NT} [233, 401] and $htt^{NT}Q_NP_{11}$ [137] monomers showed that the htt^{NT} region can populate alpha-helical configurations, while others suggest a rather amorphous monomer and dimer for $htt^{NT}Q_N$ [504]. The htt^{NT} region stabilizes β-sheet structures in the Q_N region of small oligomers of $htt^{NT}Q_N$ [110], while the P_N region, for its part, reduces the β-sheet probability of the Q_N region in the exon 1 of huntingtin [260]. Overall, combining experiments and simulations helped to understand the structural changes occurring in $htt^{NT}Q_N$ during its aggregation without yet leading to a consensual family of conformations.

As well as being an important modulator of the aggregation in solution, both the presence of the htt^{NT} region and the number of glutamines in the Q_N region also modulate the interactions of huntingtin with phospholipid membranes. In the cell, the htt^{NT} region of huntingtin strongly influences its localization on the mitochondria, endoplasmic reticulum and Golgi [16, 398]. Interestingly, htt^{NT}'s amino acids sequence shares similarities with signalling peptides as it is am-

phipathic and it displays, when mapped to the α-helix topology, an amphipathic plane separating on opposite sides its non-polar from its charged amino acids. It could therefore play a key part in huntingtin's physiological roles by controlling its localization in the cell. The length of Q_N also modulates the localization of huntingtin by increasing the nuclear entry [120, 134] leading, when above the pathological threshold, to deleterious effects on the cell through various nuclear and extra-nuclear pathways many of which involve membrane-containing structures [38]. In vitro experiments further suggest that these pathways could involve interactions with membrane as $htt^{NT}Q_N$ disrupts DMPC :POPE (1 :1) vesicles in a Q_N-length dependent manner as shown by differential scanning calorimetry experiments [232]. Similar observations were also obtained from atomic force microscopy (AFM) experiments on supported lipid membranes from brain extracts [71]. Moreover, htt^{NT} alone can cause membrane permeation of large unilamellar vesicles as shown by calcein release assays [319]. Finally, the lipid composition also modulates these interactions as huntingtin binds more strongly in the presence of higher ratios of POPE or PI phospholipids [230, 231]. Overall, these observations indicate that the interactions of huntingtin with lipid membranes might be related to its physiological and pathological roles raising the need to precisely characterize these interactions at the atomic level.

In this paper, we develop an atomistic picture of the partitioning dynamics of $htt^{NT}Q_N$ in a phospholipid membrane through all-atom, explicit solvent molecular dynamics (MD) simulations. This study is inspired by other computational simulations that have helped characterize the structures and dynamics of the membrane interactions of other well-known amyloid proteins such as amyloid-beta (Aβ) [207, 208] and islet amyloid polypeptide (IAPP) [211, 531]. More precisely, the main goals of this study are to unveil the insertion dynamics of htt^{NT} on a phospholipid bilayer and to describe the impact of the Q_N region. Moreover, we determine the key amino acids, through their interactions with the phospholipid membrane, at the origin of $htt^{NT}Q_N$ structural stability and orientation using a detailed residue-level analysis combining insertion depth, solvent accessibility, hydrogen bond and salt-bridge measurements. This information is of outmost importance as it complements previous experimental studies on $htt^{NT}Q_N$ – membrane systems for which such a level of atomic precision is missing.

Our paper is constructed as follows. We first describe the simulation parameters used, the choice of protein and phospholipid forcefields, the preparation of the initial configurations, and the analysis protocol. We next present our results on the interactions of the htt^{NT} and $htt^{NT}Q_{20}$

fragments with a phospholipid bilayer. Finally, we relate our results to previous experiments and discuss their implications for the huntingtin protein.

11.3 Materials and Methods

We investigate the atomistic origin of adsorption, insertion and stability of huntingtin N-terminal ($htt^{NT}Q_N$) on a phospholipid membrane composed of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) phospholipids using all-atom, explicit solvent molecular dynamics simulations. We also investigate the effect of the number of glutamines in Q_N by simulating fragments counting 0, 10 and 20 glutamines, with more emphasis on 0 and 20. The amino sequence of $htt^{NT}Q_N$ is MATLEKLMKAFESLKSFQ_N-NME with the subscript N referring to the number of glutamines. Since this sequence is present at the N-terminal of the longer huntingtin sequence, we N-methylated its C-terminal in order to remove the carboxylate's negative charge which could otherwise impact our results. In the current study, we choose POPE phospholipids since experiments show that the aggregation of the huntingtin N-terminal is modulated by the number of glutamines for membrane containing higher ratios of this phospholipid [230]. Moreover, experiments show that $htt^{NT}Q_N$ perturbs the physical properties of DMPC :POPE (50 :50 M ratio) membranes [232]. Finally, we choose to construct our membranes using a single phospholipid species as we want to uncouple the effect of POPE from other phospholipids.

All simulations, which are summarized in Table 11.I, are performed for at least 500 ns for a total simulation time of 15.4 μs. The simulations parameters as well as the preparation protocol for the initial configurations are described next.

MD simulations. All molecular dynamics (MD) simulations are performed with the software GROMACS version 4.5.4 [42, 185, 387, 478]. Simulations are done in the NPT ensemble using the AMBER99sb*-ILDN forcefield for the protein parameters [46, 190, 281, 382] and the Stockholm lipids (SLIPIDS) forcefield for the phospholipid parameters [203–205, 244] as justified below. We use the same simulation parameters and water model (TIP3P) as in the original SLIPIDS paper [204]. The temperature is set at 303 K using the Nosé-Hoover thermostat [188, 347] with a coupling constant of 0.5 ps. The pressure is semi-isotropically set at 1.013 bar using the Parrinello-Rahman barostat [348, 368] with a coupling constant of 10 ps. Lennard-Jones energy is switched to zero from 1.4 nm to 1.5 nm. Electrostatic interactions are treated by Particle-Mesh-Ewald (PME) [116, 150] with a real space cut-off of 1.4 nm. Bond lengths are

Table 11.I – Summary of the performed simulations.

Simulations	Time ns	Initial conf.	Box size nm	H_2O No.	POPE No.
168–POPE–hexa	350	—	$a = b = 7.3$, $c = 8.6$	6 692	168
240–POPE–hexa	350	—	$a = b = 8.8$, $c = 8.6$	9 793	240
328–POPE–hexa	350	—	$a = b = 10.4$, $c = 8.5$	13 328	328
sol–httNT	650×2	α	$a = b = c = 8.1$	13 284	0
AMBER99sb*-ILDN					
sol–httNT	500×2	α	$a = b = c = 8.1$	13 284	0
OPLS-AA					
sol–httNT	500×2	α	$a = b = c = 8.1$	13 284	0
CHARMM27					
sol–httNT	500×2	α	$a = b = c = 8.1$	13 284	0
AMBER03					
httNT_α	500×5	α	$a = b = 7.4$, $c = 11.0$	10 613	168
httNT_coil	500	coil	$a = b = 7.4$, $c = 10.1$	9 384	168
httNT_α–ins	500×2	inserted α	$a = b = 7.4$, $c = 89.9$	10 613	168
httNTQ20_α	500×5	α	$a = b = 10.3$, $c = 12.4$	24 466	328
httNTQ20_coil	500	coil	$a = b = 10.3$, $c = 11.5$	21 930	328
httNTQ20_α–ins	500×2	inserted α	$a = b = 10.3$, $c = 9.3$	10 613	168
httNTQ20_α-coil–ins	500	inserted α-coil	$a = b = 10.4$, $c = 9.2$	10 613	168
httNTQ10_α	500×2	α	$a = b = 8.8$, $c = 11.1$	15 462	240
httNTQ10_coil	500	coil	$a = b = 8.8$, $c = 11.2$	15 697	240

The box sizes correspond to the initial length of the principal axes which are subject to small variation during the simulations as they are done in the NPT ensemble. The periodic boxes are either an octahedron ($\alpha = 70.5°$, $\beta = 109.5°$, $\gamma = 70.5°$) for the simulations in solution or hexagonal prism ($\alpha = 90°$, $\beta = 90°$, $\gamma = 120°$) for the simulations with a phospholipid membrane. All simulations are done using the AMBER99sb*-ILDN forcefield for the protein unless otherwise specified.

constrained using LINCS [184] and water geometry with SETTLE [326] allowing an integration time step of 2 fs. The center of mass linear momentum is removed every 20 ps. Unless otherwise stated, all these parameters are also used to prepare and equilibrate the initial configurations of the protein, membrane and protein-membrane systems.

Membranes are prepared as follow. We build membranes of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) with a hexagonal cross-section from the original 128 POPE rectangular membrane presented in the original SLIPIDS paper [204] (http ://people.su.se/~jjm/). Three different size of hexagonal membranes are created : 168 phospholipids for htt^{NT}, 240 phospholipids for $htt^{NT}Q_{10}$ and 328 phospholipids for $htt^{NT}Q_{20}$. Each system size is chosen such that the distances between the peptide and its periodic images are always greater than 3.0

nm during all simulations. Each membrane is then equilibrated for 350 ns. The areas per lipid averaged over the last 100 ns are 0.563 ± 0.009 nm^2 (168–POPE–hexa), 0.562 ± 0.007 nm^2 (240–POPE–hexa) and 0.559 ± 0.007 nm^2 (328–POPE–hexa), in agreement with the *in silico* value of 0.562 ± 0.004 nm^2 obtained in the original SLIPIDS paper [204] and with the experimental value [391] of 0.566 nm^2. The order parameters of the acyl chains averaged over the last 100 ns also agree with the original SLIPIDS paper [204] as shown in Figure IV.1.

Using the equilibrated membranes, the initial protein-membrane systems are prepared as follow. Two different initial configurations are used for each peptide : a single α-helix and a single random coil (Figure IV.2). The random coil configuration is generated from a 10-ns MD simulation in solution starting from a completely extended structure. For both initial configurations, the protein center of mass is placed at 4.5 nm from the membrane center-of-mass, and the principal axis of the backbone atoms is aligned parallel to the membrane such that the initial minimum distance between the protein and the membrane is greater than 2.0 nm, that is the protein is at least 0.5 nm beyond real space cutoffs from the membrane. Finally, the box is filled with water molecules and two chloride ions are added to neutralize each system. A pre-inserted α-helix is also used as initial configuration for htt^{NT} and htt^{NT}Q$_{20}$ (Figure IV.2). Using InflateGRO2 [413], the peptide is inserted such that the principal axis of its backbone atoms is aligned parallel to the membrane surface and that its non-polar residues are pointing toward the hydrophobic core of the membrane. For htt^{NT}Q$_{20}$, a configuration with an α-helical htt^{NT} region and a random coiled Q$_N$ region is also used. This initial state is obtained from a 10-ns MD simulation starting from the pre-inserted and equilibrated single α-helix configuration during which the heavy atoms of the htt^{NT} region are harmonically restrained to their initial position and the ψ dihedral angles of one each two residues in the Q$_N$ region are restrained harmonically to 135°. These restrains allow to keep the htt^{NT} region folded into an α-helix as well as disordering the Q$_N$ region during the preparation of this initial state.

Prior to each MD simulation, we equilibrate the initial configuration by first minimizing the energy using the conjugate gradient algorithm with a steepest descent step applied once every 100 steps, and by secondly equilibrating the system for 1 ns using a 1 fs time step and restrains on all heavy atoms of the protein. The 0 ns configuration in all figures refers to the structures obtained after these pre-MD steps.

Forcefield. The forcefield selection is motivated by several aspects, the most important of

which is that the protein forcefield correctly samples the equilibrium secondary structure of huntingtin 17-amino-acid N-terminal (htt^{NT}) positioned just before the consecutive segment of glutamines (Q_N). Circular dichroism (CD) spectroscopy studies show that the htt^{NT} peptide samples some α-helical content in solution [16, 319, 459, 504]. Nuclear magnetic resonance (NMR) studies further show that there is no stable α-helix in solution [459]. We evaluate different forcefields by simulating this peptide in solution using the simulation parameters previously presented. As described in the Supplementary Material, some popular protein forcefields either overestimate (CHARMM27 [299, 300] and AMBER03 [141]) or underestimate (OPLS-AA\L [217, 220]) the fraction of α-helices sampled by htt^{NT}. On the other hand, we found that AMBER99sb*-ILDN [382] has the least bias toward the fully formed alpha-helix or the fully random-coil states, two ensembles that do not correspond to the aforementioned CD and NMR experimental results on the htt^{NT} peptide in solution. Moreover, this latter forcefield correctly reproduces the folding and stability of diverse proteins as shown by microsecond time scale MD simulations [28, 98, 279, 383]. Therefore, for our simulations, we select the forcefield AMBER99sb*-ILDN which originates from the combination of two corrections to AMBER99sb [190] : AMBER99sb-ILDN that corrects the energy landscape of the χ_1 and χ_2 torsion angles of the isoleucine, leucine, aspartate and asparagine residues [281], and AMBER99sb* that corrects the α-helix–coil imbalance impacting the conformational sampling [46].

Once we determined that AMBER99sb*-ILDN best matches the structural ensemble of the htt^{NT} peptide in solution, we choose a phospholipid forcefield that is compatible with the AMBER forcefields family, usable in the NPT ensemble as well as good at reproducing diverse bilayer properties measured experimentally. As previously shown, the Stockholm lipids (SLIPIDS) satisfy these three conditions on time scales of hundreds of nanoseconds for many phospholipids [203–206]. The combination SLIPIDS and AMBER has been used to compute the free energy of insertion of several compounds into phospholipid bilayers [206] and the tilt angle of the WALP23 peptide inserted in phospholipid bilayers [204]. The computed quantities are in good agreement with experimental measurements [204, 206].

Analysis. To characterize the $htt^{NT}Q_N$ – membrane system, we compute several quantities using in-house codes and GROMACS utilities, unless otherwise stated. To describe the backbone structure, the α-helix propensity is computed using STRIDE [164]. To quantify the degree of insertion, the center-of-mass distance between each residue of $htt^{NT}Q_N$ and the membrane, as

well as the solvent accessible surface area using a water probe radius of 1.4 Å [146] are computed. To quantify the membrane properties, the order parameters of the acyl chains are computed using the order program of the MDynaMix software [295]. To quantify the peptide–membrane interactions, salt-bridges are considered when two oppositely charged groups are within 4 Å of each other [21], and hydrogen bonds are considered when the donor–acceptor distance is smaller than 3.5 Å and the hydrogen–donor–acceptor angle is smaller than 30°. To cluster the configurations, the root-mean square distance (RMSD) between the backbone atoms is first computed on any pair of configurations which are then regrouped into clusters from the largest to the smallest by removing the clustered configurations each time a new cluster is found. The clusters have the following properties : their center has a backbone RMSD of at most 2.0 Å against all other configurations inside the cluster and a given configuration can not be found in two different clusters. All figures showing the time-evolution of some quantity display the running time average of the data using 5-ns time windows. All reported error bars on averaged quantities correspond to one statistical standard deviation. Finally, molecular graphics images are generated using the PyMOL software (http ://www.pymol.org).

11.4 Results

11.4.1 htt^{NT} with a bilayer

To investigate the insertion dynamics of htt^{NT} alone, we launch five independent MD simulations starting from the same initial configuration with different initial velocity distributions (11.I). We choose to start with a single α-helix state for two main reasons : (1) htt^{NT} samples widely different α-helical configurations in solution with relatively similar weights (Figures IV.3 and IV.4), so choosing a specific one over another might induce a bias and (2) the atomic structure of htt^{NT} alone in solution has not yet been resolved by NMR or X-rays crystallography. As shown below, this particular choice does not seem to bias our results since similar results are obtained when a random coil configuration is taken for the initial state.

Three trajectories show partial insertion of the non-polar residues of the htt^{NT} peptide and they are displayed in Figure 11.1 (simulations httNT_α_1 and httNT_α_2) and Figure IV.5 (simulation httNT_α_3). These are characterized by four main steps : (1) an approach step that seems to be driven by long-range electrostatic interactions between the peptide and the phospholipids' head group, (2) a reorganization step during which the peptide undergoes structural

changes in order to bring the non-polar residues nearer to the membrane core, (3) an anchoring step initialized by the positioning of a non-polar residue such as the phenylalanines inside the hydrophobic core of the membrane, and (4) an insertion step during which the other non-polar residues are partitioned in the membrane as structural reorganization continues.

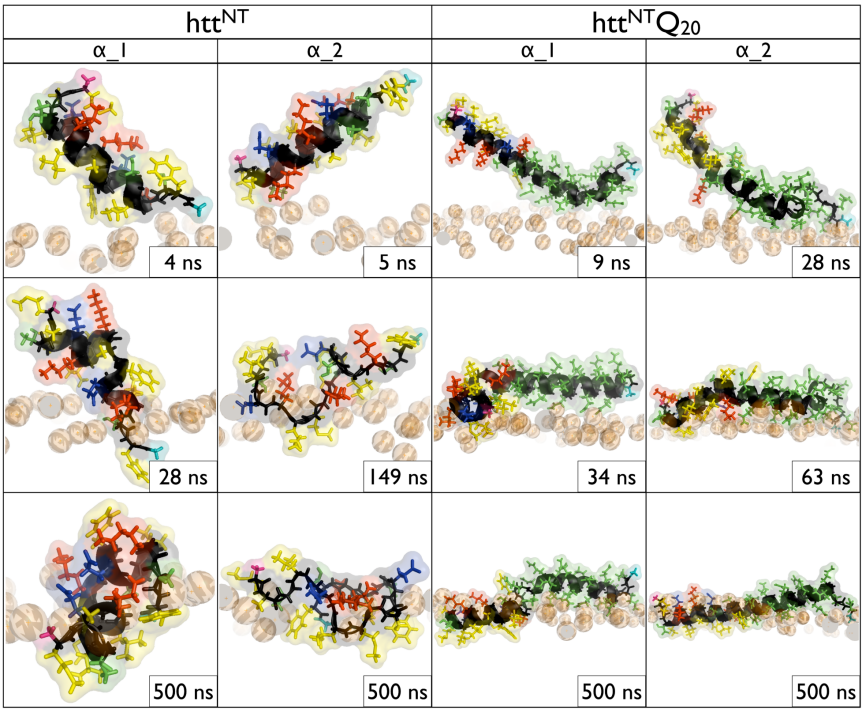


Figure 11.1 – The httNT_α_1 and httNT_α_2 trajectories for htt^{NT} are shown on the first and second columns, respectively. The httNTQ20_α_1 and httNTQ20_α_2 trajectories for htt^{NT}Q$_{20}$ are shown on the third and fourth columns, respectively. The non-polar, negatively charged, positively charged and polar amino acids of the peptide are respectively shown in yellow, red, blue and green. The backbone atoms are displayed in black, and the N- and C-terminal are respectively shown in pink and teal. Only the phosphorus atom is displayed as an orange sphere for the phospholipids.

Because of the large accessible configurational space, considerable differences occur in the details of each step. These differences are summarized in Table 11.II which shows, in order of occurrence, the residues forming their first contact with the bilayer for the approach step as well as those found below the phosphate groups of the bilayer for the anchoring and insertion steps. First, the approach step seems to be initiated by different charged parts of the peptide : Lys-15 is within the first residues to come into contact with the membrane for the httNT_α_1 and httNT_α_3 trajectories, while it is the N-terminal for httNT_α_2. Second, structural reorganization can be quite different as the htt^{NT} peptide stays mostly α-helical in the first trajectory (55

$\pm$ 18%), while it becomes largely disordered in the second and third trajectories (0 $\pm$ 1% and 2 $\pm$ 8%, respectively) as shown in Figure 11.2. Third, the anchoring step can be initiated by a phenylalanine (Phe-17 for httNT_α_1 and Phe-11 for httNT_α_3) or by other non-polar residues (Leu-7 for httNT_α_2). Fourth, the insertion step can occur in a different order of residues : Leu^{7}–Met^{8}–Leu^{4}–Met^{1} for httNT_α_1, Met^{8}–Phe^{11}–Leu^{14}–Phe^{14} for httNT_α_2, and Leu^{7}–Met^{8}–Met^{1}–Leu^{4} for httNT_α_3. Finally, the httNT_α_1 trajectory seems to have reached a metastable state with many inserted non-polar residues (Figure 11.3). For their part, the other two trajectories are less inserted and still undergo small structural changes as their non-polar residues sample a wider range of positions from the bilayer's center.

The last two trajectories (simulations httNT_α_4 and httNT_α_5), for which the peptide does not insert and remains trapped between the third and fourth steps, are shown in Figure IV.5. Looking more closely, the peptide's attempts to bring its non-polar residues toward the hydrophobic core of the membrane are unsuccessful because it needs to undergo structural changes that are impaired by strong electrostatic interactions between its charged residues and the phospholipids' head group (Figure IV.6). As a result, these trajectories are significantly less inserted and still undergo structural reorganizations (Figure IV.7 vs. Figure 11.3). Interestingly, these results suggest that the amount of α-helical content does not automatically correlate with a faster insertion of the non-polar residues as these two trajectories, which do not insert, have a higher α-helical propensity than the httNT_α_2 and httNT_α_3 trajectories that insert (Table 11.II). It is expected, however, from the htt^{NT}'s sequence, that insertion of all non-polar residues requires the peptide to be in an α-helix.

We also investigate the insertion dynamics of htt^{NT} starting from a random coil that is not interacting with the membrane. Partial insertion of the non-polar residues is observed in that trajectory which follows the same four main steps observed for the trajectories obtained when starting from a single α-helix conformation (Table 11.II and Figure IV.5, simulation httNT_coil). The approach and anchoring steps are respectively initialized by the N-terminal and Phe-17, and the α-helical content of the peptide does not increase significantly during the simulation. This further confirms that a significant α-helical propensity is not a necessary condition for a partial insertion of the non-polar residues of htt^{NT} into the membrane. Further reorganization would, of course, be expected to occur if given enough time.

Overall, the completion of the four steps can span very different time scales from a several

Table 11.II – Insertion steps for all simulations done on htt^{NT}, $htt^{NT}Q_{10}$ and $htt^{NT}Q_{20}$.

Simulations	Approach step	Anchoring step	Insertion step	Average α-helix (%)
httNT_α_1	Phe^{17} Lys^{15} Ser^{16} Leu^{14} Phe^{11}	Phe^{17}	Leu^{7} Met^{8} Leu^{4} Met^{1}	55 ± 18
httNT_α_2	Met^{1} Leu^{4} Ala^{2} Glu^{5} Met^{8} Thr^{3}	Leu^{7}	Met^{8} Phe^{11} Leu^{14} Phe^{17}	0 ± 1
httNT_α_3	Phe^{17} Phe^{11} Lys^{15} Ser^{16} Met^{8} Leu^{4}	Phe^{11}	Leu^{7} Met^{8} Met^{1} Leu^{4}	2 ± 8
httNT_α_4	Phe^{17} Lys^{15} Ser^{16} Phe^{11} Met^{8} Glu^{12}	Phe^{17}	–	58 ± 22
httNT_α_5	Glu^{12} Lys^{15} Ser^{16} Ser^{13} Ala^{10} Phe^{17}	Phe^{17}	–	34 ± 16
httNT_coil	Met^{1} Ala^{2} Leu^{4}	Leu^{4}	Phe^{17}	7 ± 12
httNT_α–ins_1	–	–	–	76 ± 15
httNT_α–ins_2	–	–	–	83 ± 8
httNTQ20_α_1	Q_N Lys^{9} Lys^{6}	Leu^{4}	Met^{1} Met^{8} Phe^{11} Leu^{7} Leu^{14}	72 ± 8
httNTQ20_α_2	Q_N Phe^{17}	Leu^{4}	Met^{8} Leu^{7} Phe^{11} Leu^{14} Phe^{17}	76 ± 11
httNTQ20_α_3	Q_N	–	–	64 ± 3
httNTQ20_α_4	Q_N Lys^{9}	Met^{1}	–	31 ± 21
httNTQ20_α_5	Q_N	–	–	80 ± 9
httNTQ20_coil	Q_N	–	–	2 ± 7
httNTQ20_α–ins_1	–	–	–	84 ± 8
httNTQ20_α–ins_2	–	–	–	81 ± 7
httNTQ20_α-coil–ins	–	–	–	74 ± 13
httNTQ10_α_1	Q_N	Met^{1}	–	42 ± 13
httNTQ10_α_2	Q_N	–	–	0 ± 4
httNTQ10_coil	Q_N	–	–	8 ± 13

The second column (approach step) displays the amino acids that contact the phospholipid membrane within 1 ns after the occurrence of the first contact. The third column (anchoring step) indicates the first non-polar amino acid to be inserted in the membrane. The fourth column (insertion step) shows the non-polar amino acid that partition in the membrane in order of occurrence for those that stay inserted up to the end of the simulation. The last two columns respectively displays the average α-helix propensity only on the htt^{NT} region over the last 200 ns. For the initially inserted peptide trajectories, the average is taken over the last 300 ns.

tenths to hundreds of nanoseconds and beyond our computational capacity. A bottleneck seems to be particularly difficult to overcome using standard MD simulations on the sub-microsecond time scale. Insertion of all non-polar residues requires important structural changes that can be

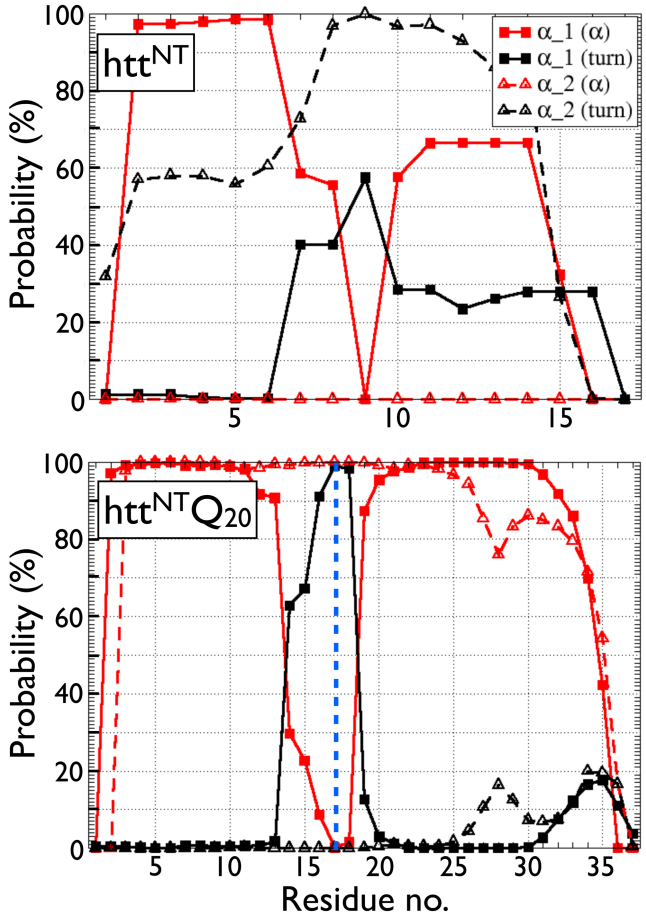


Figure 11.2 – The per residue secondary structure propensities for the trajectories of htt^{NT} for httNT_α_1 and httNT_α_2 (top) and $htt^{NT}Q_{20}$ for httNTQ20_α_1 and httNTQ20_α_2 (bottom). The average propensities of α-helix and turn are respectively shown in red and black for the first (solid lines, solid squares) and the second trajectory (dotted lines, empty triangles). The average secondary structure is computed over the last 200 ns of the simulations. The blue dotted line on the bottom panel indicates the last amino acid of the htt^{NT} region (Phe^{17}).

impaired by the strong electrostatic interactions between the charged residues of the peptide and the phospholipids' head group as these must be temporarily broken in order to reach a more inserted state. As a result, large fluctuations occur in the Coulomb energy between the peptide and membrane, while the Lennard-Jones energy undergoes significantly smaller fluctuations (Figure IV.6).

Having said that, it is nonetheless important to have an idea of what to expect for the final state of htt^{NT}. In order to address this, we investigate the stability of an initially inserted htt^{NT} starting from a well-structured single α-helix configuration (simulations httNT_α–ins_1 and

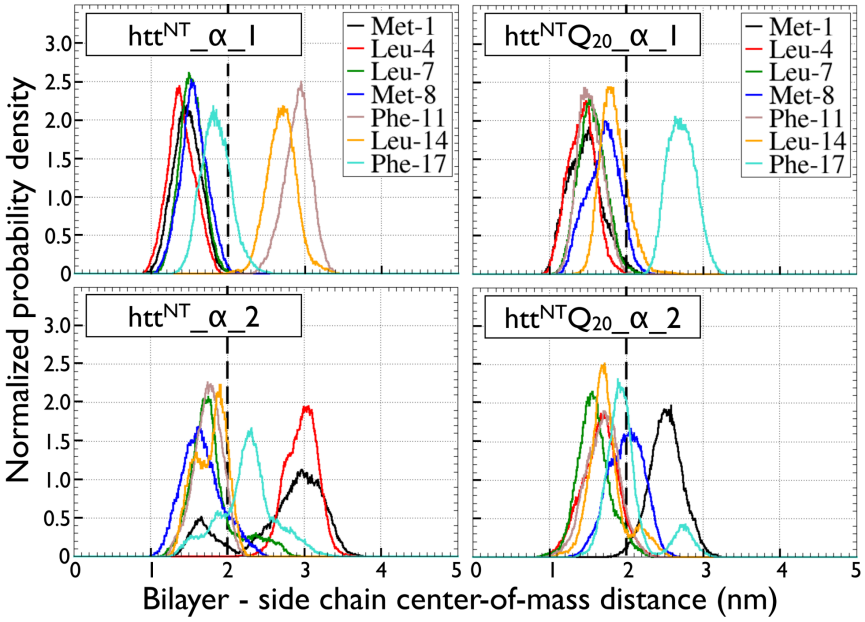


Figure 11.3 – Position probability densities of all non-polar side-chains with respect to the center-of-mass of the phospholipid bilayer of htt^{NT} for httNT_α_1 and httNT_α_2 (left) and htt^{NT}Q$_{20}$ for httNTQ20_α_1 and httNTQ20_α_2 (right). The average position of the phospholipids' phosphate group is represented by the thick black dotted line near 2.0 Å. The average is taken over on the 300-500 ns time interval.

httNT_α–ins_2) with initial tilt / rotational pitch angles of 86° / 93°. Two independent 500 ns simulations, with different initial velocity distributions, show that the α-helical state is stable when the non-polar residues of htt^{NT} are oriented toward the hydrophobic core of the membrane (Figure 11.4).

Over the last 300 ns, the α-helix is mainly parallel to the membrane with tilt / rotational pitch angles of 89 $\pm$ 8° / 114 $\pm$ 11° (httNT_α–ins_1) and 89 $\pm$ 9° / 100 $\pm$ 14° (httNT_α–ins_2) as shown in Table 11.III. The α-helical propensities are 76 $\pm$ 15% (httNT_α–ins_1) and 83 $\pm$ 8% (httNT_α–ins_2) with a stable α-helix from Ala-2 to Ser-16 (Figure 11.5). The structure and orientation of htt^{NT} are stabilized by key peptide–membrane interactions through salt-bridges involving Glu-5, Glu-12, Lys-6 and Lys-15, as well as hydrogen bonds involving Thr-3 and Ser-13 (Table 11.IV). Non-polar residues, mostly Leu-4, Leu-7, Phe-11 and Leu-14, are positioned away from the solvent in the hydrophobic core of the membrane (Figure 11.6). We therefore expect that the htt^{NT} peptide in our previous simulations (Figure 11.1 and Figure IV.5) could undergo further structural reorganizations toward the stable α-helical configuration shown in Figure 11.4 given sufficient time.

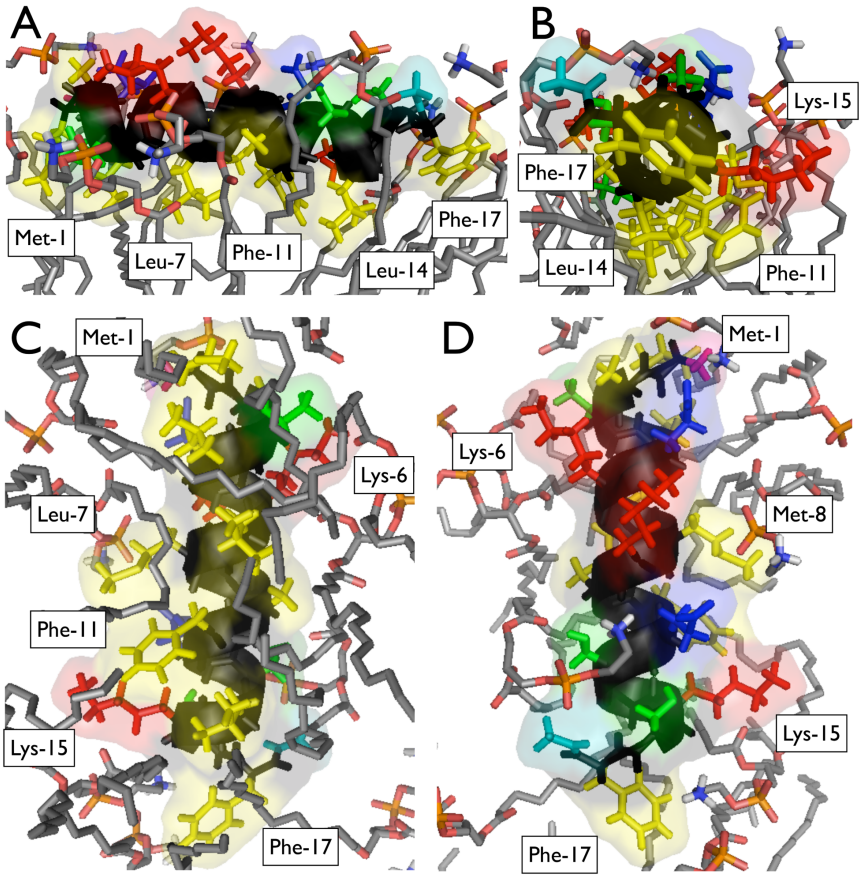


Figure 11.4 – Atomic structure of the average orientation of htt^{NT} over the last 300 ns of httNT_α–ins_1. (A) Side view, (B) view from the C-terminal, (C) view from the membrane and (D) view from the solvent. The non-polar, negatively charged, positively charged and polar amino acids of the peptide are respectively shown in yellow, red, blue and green. The backbone atoms are displayed in black, and the N- and C-terminal are respectively shown in pink and teal. The phosphorus, oxygen, nitrogen and carbon atoms of the phospholipids are respectively displayed in orange, red, blue and grey.

11.4.2 htt^{NT}Q$_{20}$ with a bilayer

We now investigate the effect of the polyglutamine segment (Q$_N$) on the dynamics and equilibrium properties of the htt^{NT}Q$_N$ – phospholipid bilayer system. Five independent 500 ns molecular dynamics simulations each starting from the same initial structure with different initial velocity distributions were executed (Table 11.I). As for the htt^{NT} peptide, the initial structure is a single α-helix since there is no high-resolution experimental structure of htt^{NT}Q$_N$ in solution at the moment. This choice does not seem to impact the sampling since very different trajectories are observed showing either insertion of almost all the non-polar residues of the htt^{NT} region (for 2 trajectories) or no insertion at all (for 3 trajectories).

Table 11.III – Orientation of the amphipathic plane of the htt^{NT} region.

Simulation	Tilt angle	Pitch angle	RMSD (Å)
httNT_α–ins_1	$89 \pm 8°$	$114 \pm 11°$	2.3 ± 0.2
httNT_α–ins_2	$89 \pm 8°$	$100 \pm 14°$	2.3 ± 0.2
httNTQ20_α–ins_1	$82 \pm 8°$	$107 \pm 12°$	2.1 ± 0.1
httNTQ20_α–ins_2	$84 \pm 9°$	$91 \pm 14°$	2.2 ± 0.2
httNTQ20_α-coil–ins	$83 \pm 6°$	$95 \pm 12°$	2.4 ± 0.4
Experimental values	$103 \pm 5°$	$137 \pm 5°$	–

Orientation of the amphipathic plane of the htt^{NT} region for simulations of both htt^{NT} and htt^{NT}Q$_{20}$ starting from an initially inserted peptide, and all-atom RMSD with respect to structured segment (residues 5 to 17) of the NMR model [318]. The tilt and rotational pitch angles are defined by the rotation of the amphipathic plane, initially in the z-x plane, around the negative z-axis (tilt) followed by the rotation around the new negative y-axis (rotational pitch) to its current orientation. The graphical definition of these angles along with the distributions of angles sampled during our simulations are shown in Figure IV.8. The experimental values have been taken from a NMR experimental study on the htt^{NT} – bilayer system [318]. In our simulations, the initial values for the tilt / rotational pitch angles are 86° / 93° (httNT_α–ins_1 and httNT_α–ins_2), 81° / 124° (httNTQ20_α–ins_1 and httNTQ20_α–ins_2) and 82° / 126° (httNTQ20_α-coil–ins). The averages are taken over the 200-500 ns time interval.

Analysis of the two trajectories characterized by the insertion of almost all the non-polar residues (simulations httNTQ20_α_1 and httNTQ20_α_2) reveals that the four mains steps observed for the htt^{NT} peptide – approach, reorganization, anchoring and insertion – also describe the dynamics of insertion in the presence of the Q$_N$ region (Figure 11.1). It is in the details of each step that differences, mainly due to Q$_N$, are revealed as summarized in Table 11.II. For instance, the lysines, as for the htt^{NT} peptide, but also the carbonyl and amine groups of the glutamines side-chain in the Q$_N$ region can be within the first to interact with the membrane. Furthermore, during the reorganization step, these glutamines interact strongly with the phospholipids' head group through hydrogen bonds and the structural changes are less important than for htt^{NT} (Table 11.II). More precisely, the largest structural changes are observed in the httNTQ20_α_1 trajectory between residues Leu-14 and Phe-17 that adopt a turn conformation linking the inserted α-helical htt^{NT} to the adsorbed Q$_N$ (Figure 11.2). By interacting strongly with the phospholipids' head group, the Q$_N$ region then provides a stable scaffold for the partitioning of the non-polar residues of the htt^{NT} region inside the hydrophobic core of the membrane. As a result, htt^{NT}Q$_{20}$ is more structured and more inserted than htt^{NT}. Over the last 200 ns, the position density of each non-polar side-chain shows that the peptide has reached a metastable

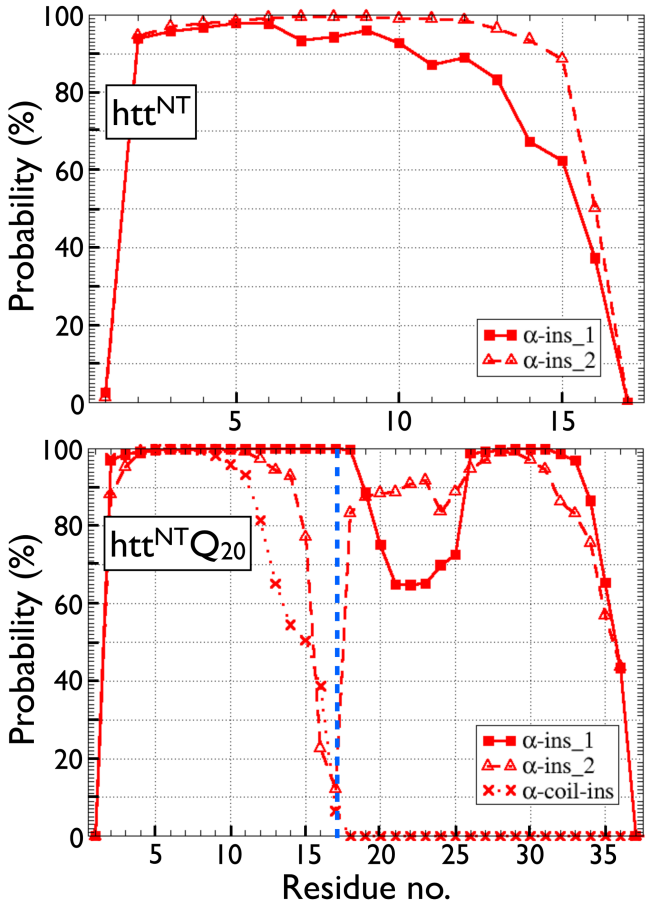


Figure 11.5 – Same as Figure 11.2, for the simulations starting from an initially inserted peptide for htt^{NT} (httNT_α–ins_1 and httNT_α–ins_2) and $htt^{NT}Q_{20}$ (httNTQ20_α–ins_1, httNTQ20_α–ins_2 and httNTQ20_α-coil–ins). The average is taken over the 200-500 ns time interval.

state with a more pronounced insertion than for the htt^{NT} peptide (Figure 11.3).

The insertion of $htt^{NT}Q_{20}$ does not occur for the other three trajectories (simulations httNTQ20_α_3, httNTQ20_α_4 and httNTQ20_α_5 displayed in Figure IV.9) indicating that this process is not necessarily faster in the presence of Q_N. Similarly to htt^{NT}, these trajectories are trapped in the reorganization step (second step) due to the strong electrostatic interactions between the charged residues and the phospholipids' head group (Figure IV.6). As a result, the non-polar residues are trapped on the opposite side of the membrane and cannot be inserted in its hydrophobic core. Trajectories httNTQ20_α_3 and httNTQ20_α_5 undergo few structural changes during the last 200 ns as the positions of their non-polar residues do not change much, while the position densities are broader in httNTQ20_α_4 suggesting that this trajectory is not completely trapped (Figure IV.10).

Table 11.IV – Per residue htt^{NT}–membrane interactions of httNT_α–ins_1 and httNT_α–ins_2.

Simulations	Amino acids	Insertion depth (nm)	SASA (nm^2)	H-bonds (no.)	Salt-bridges (no.)
httNT_α–ins_1	htt^{NT}	1.9 ± 0.2	5.0 ± 0.9	3 ± 1 (98%)	2 ± 1 (80%)
	Met-1	1.6 ± 0.2	0.03 ± 0.06	–	–
	Ala-2	2.2 ± 0.2	0.4 ± 0.1	–	–
	Thr-3	1.7 ± 0.2	0.06 ± 0.07	1.0 ± 0.2 (72%)	–
	Leu-4	1.5 ± 0.2	0.01 ± 0.03	–	–
	Glu-5	2.2 ± 0.2	0.7 ± 0.3	1.1 ± 0.3 (30%)	1.0 ± 0.2 (31%)
	Lys-6	2.1 ± 0.2	0.7 ± 0.3	1.2 ± 0.5 (41%)	1.2 ± 0.4 (31%)
	Leu-7	1.4 ± 0.2	0.01 ± 0.02	–	–
	Met-8	1.8 ± 0.2	0.1 ± 0.1	–	–
	Lys-9	2.4 ± 0.2	1.1 ± 0.2	1.0 ± 0.2 (7%)	1.0 ± 0.2 (7%)
	Ala-10	1.8 ± 0.2	0.03 ± 0.04	–	–
	Phe-11	1.5 ± 0.2	0.01 ± 0.02	–	–
	Glu-12	2.2 ± 0.2	0.7 ± 0.2	1.0 ± 0.2 (39%)	1.1 ± 0.1 (40%)
	Ser-13	2.1 ± 0.2	0.2 ± 0.1	1.0 ± 0.1 (24%)	–
	Leu-14	1.5 ± 0.2	0.01 ± 0.02	–	–
	Lys-15	1.9 ± 0.2	0.3 ± 0.2	1.1 ± 0.4 (58%)	1.1 ± 0.3 (43%)
	Ser-16	2.2 ± 0.3	0.4 ± 0.1	1.0 ± 0.1 (11%)	–
	Phe-17	1.7 ± 0.3	0.1 ± 0.2	–	–
httNT_α–ins_2	htt^{NT}	1.9 ± 0.2	5.0 ± 0.8	3 ± 1 (100%)	2 ± 1 (84%)
	Met-1	1.7 ± 0.2	0.04 ± 0.07	–	–
	Ala-2	2.2 ± 0.2	0.3 ± 0.1	–	–
	Thr-3	1.7 ± 0.2	0.03 ± 0.05	1.0 ± 0.1 (89%)	–
	Leu-4	1.6 ± 0.2	0.01 ± 0.01	–	–
	Glu-5	2.3 ± 0.2	0.7 ± 0.2	1.0 ± 0.1 (35%)	1.0 ± 0.1 (36%)
	Lys-6	2.1 ± 0.2	0.6 ± 0.2	1.3 ± 0.5 (57%)	1.2 ± 0.4 (46%)
	Leu-7	1.5 ± 0.2	0.01 ± 0.01	–	–
	Met-8	2.0 ± 0.2	0.3 ± 0.1	–	–
	Lys-9	2.4 ± 0.2	1.1 ± 0.2	1.0 ± 0.2 (6%)	1.0 ± 0.1 (5%)
	Ala-10	1.7 ± 0.2	0.01 ± 0.02	–	–
	Phe-11	1.6 ± 0.2	0.02 ± 0.03	–	–
	Glu-12	2.3 ± 0.2	0.8 ± 0.2	1.2 ± 0.5 (17%)	1.1 ± 0.3 (18%)
	Ser-13	2.1 ± 0.2	0.2 ± 0.1	1.0 ± 0.1 (47%)	–
	Leu-14	1.5 ± 0.2	0.01 ± 0.01	–	–
	Lys-15	1.9 ± 0.2	0.3 ± 0.2	1.2 ± 0.4 (57%)	1.1 ± 0.3 (48%)
	Ser-16	2.2 ± 0.2	0.4 ± 0.1	1.0 ± 0.1 (5%)	–
	Phe-17	1.7 ± 0.4	0.2 ± 0.3	–	–

The degree of insertion of each residue in the membrane is characterized by their insertion depth, which is defined as the center-of-mass distance between the side chain and the membrane, and their solvent accessible surface area (SASA) as shown on the third and fourth columns respectively. For comparison, the upper layer of phosphorous is at 2.0 nm from the center of the membrane. The number of salt-bridges and hydrogen bonds of each residue with the phospholipids are shown on the fifth and sixth columns respectively. The average for these latter two quantities is taken over non-zero values only whose occurrence is shown in parenthesis. Averages are taken over the last 300 ns of each trajectory.

We also investigate the dynamics of htt^{NT}Q$_{20}$ insertion starting from a random coil (simulation httNTQ20_coil displayed in Figure IV.9). Contrary to what is observed with the peptide without the Q_N segment, non-polar residues are not found inside the hydrophobic core of the membrane and the peptide adopts rather a random coil globular form. Few structural changes occur during the reorganization step (Figure IV.10) because the charged residues and the gluta-

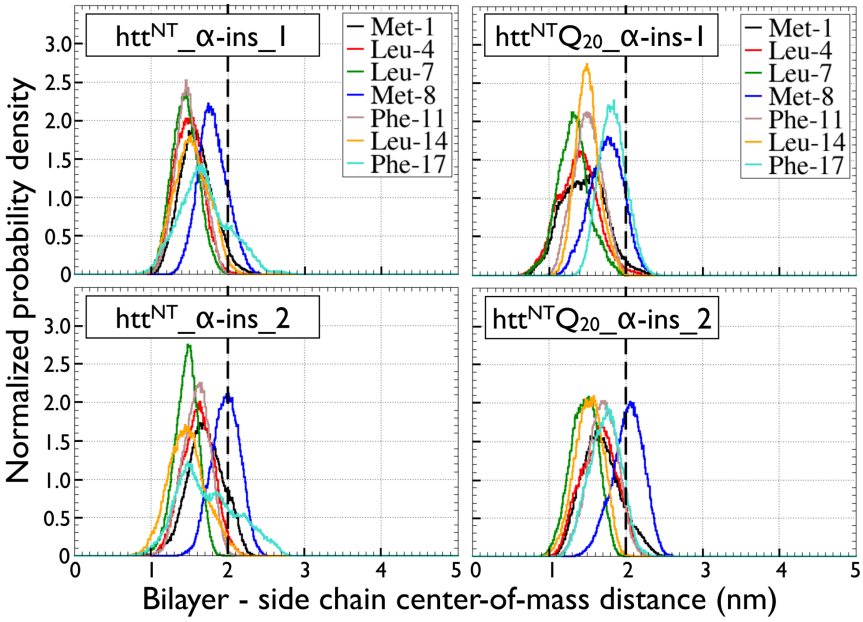


Figure 11.6 – Same as Figure 11.3, for the simulations starting from an initially inserted peptide for htt^{NT} (httNT_α–ins_1 and httNT_α–ins_2) and $htt^{NT}Q_{20}$ (httNTQ20_α–ins_1 and httNTQ20_α–ins_2). The average is taken over the 200-500 ns time interval.

mines interact strongly with the membrane. It is possible that the addition of the Q_N region also increases the conformational entropy of the peptide slowing down significantly the reorganization dynamics that would lead to a structurally ordered partitioning.

As for the htt^{NT} peptide, the non-inserted trajectories of $htt^{NT}Q_{20}$ are expected to undergo more structural changes on longer time scales. To investigate the possible conformations for the inserted peptide, a set of three simulations starting from an initially inserted peptide are executed. Two initial configurations are used in which $htt^{NT}Q_{20}$ is either a single α-helix (simulations httNTQ20_α–ins_1 and httNTQ20_α–ins_2) or a α/coil htt^{NT}/Q_N (simulation httNTQ20_α-coil–ins). The average orientation of the htt^{NT} region on the membrane for httNTQ20_α–ins_1 and httNTQ20_α-coil–ins are respectively shown in Figure 11.7 and Figure IV.11.

Clearly, the single α-helix structure in the htt^{NT} region is stable (Figure 11.5). The tilt and rotational pitch angles are also relatively close to the values found for htt^{NT}, although we observe a shift towards smaller tilt and pitch angles with respect to the simulations on htt^{NT} (Table 11.III). These observations do not significantly change when starting from a disordered Q_N region suggesting that its conformation plays a secondary role in the structure and orientation of the htt^{NT} region on the bilayer (Figure 11.5 and Table 11.III). These similarities between htt^{NT}

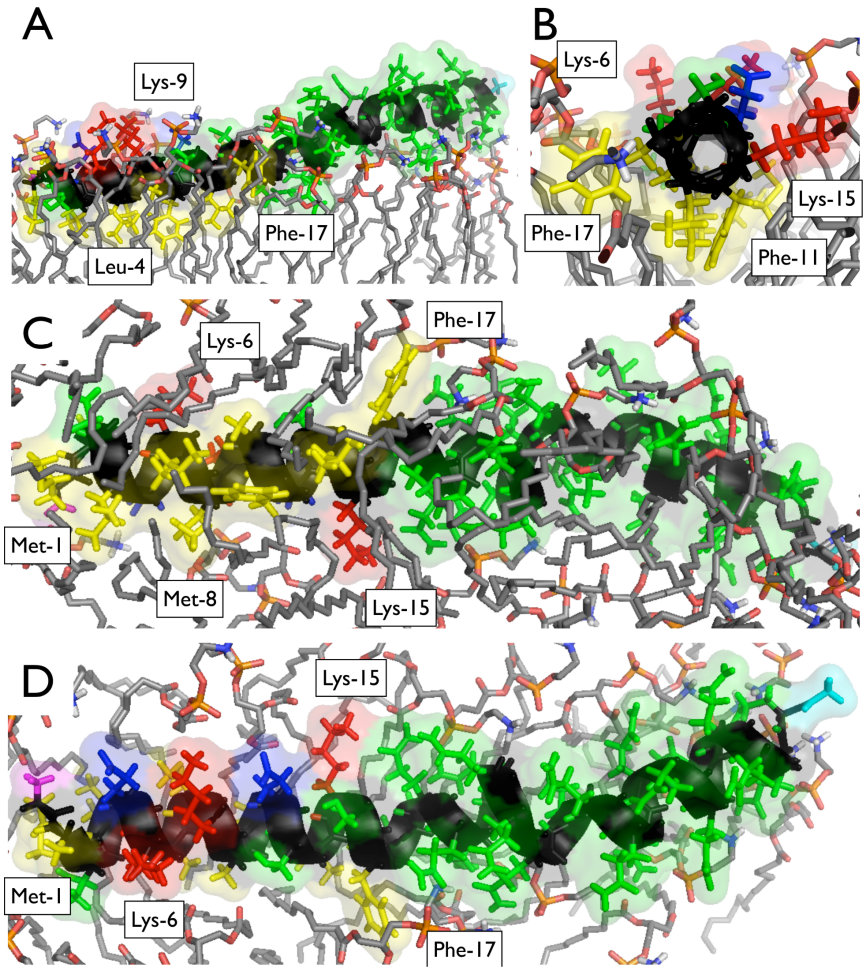


Figure 11.7 – Same as Figure 11.4, for $htt^{NT}Q_{20}$ (httNTQ20_α–ins_1).

and $htt^{NT}Q_{20}$ originate from comparable peptide–membrane interactions such as salt-bridges involving mostly Glu-5, Glu-12, Lys-6 and Lys15, hydrogen bonds involving mostly Thr-3 and Ser-13, and burying of non-polar residues involving mostly Leu-4, Leu-7, Phe-11 and Leu-14 as shown by comparing Table 11.IV with Table 11.V. These key interactions seem to be relatively independent of the conformation visited by the Q_N region as seen by comparing Table 11.V with Table 11.VI.

Table 11.V – Per residue htt^{NT}Q20–membrane interactions of httNTQ20_α–ins_1 and httNTQ20_α–ins_2.

Simulations	Amino acids	Insertion depth (nm)	SASA (nm^2)	H-bonds (no.)	Salt-bridges (no.)
httNTQ20_α–ins_1	htt^{NT}Q$_{20}$	2.2 ± 0.1	14 ± 1	13 ± 3 (100%)	2 ± 1 (90%)
	htt^{NT}	1.8 ± 0.2	4.2 ± 0.7	3 ± 1 (97%)	2 ± 1 (90%)
	Q$_{20}$	2.5 ± 0.2	10 ± 1	10 ± 3 (100%)	–
	Met-1	1.5 ± 0.3	0.02 ± 0.05	–	–
	Ala-2	2.0 ± 0.3	0.3 ± 0.2	–	–
	Thr-3	1.5 ± 0.3	0.04 ± 0.07	1.0 ± 0.1 (15%)	–
	Leu-4	1.4 ± 0.3	0.01 ± 0.02	–	–
	Glu-5	2.1 ± 0.3	0.6 ± 0.3	1.3 ± 0.5 (54%)	1.1 ± 0.3 (55%)
	Lys-6	2.0 ± 0.3	0.5 ± 0.2	1.4 ± 0.6 (56%)	1.3 ± 0.4 (40%)
	Leu-7	1.3 ± 0.2	0.01 ± 0.01	–	–
	Met-8	1.7 ± 0.2	0.2 ± 0.1	–	–
	Lys-9	2.3 ± 0.2	1.0 ± 0.2	1.0 ± 0.1 (10%)	1.0 ± 0.1 (8%)
	Ala-10	1.7 ± 0.2	0.01 ± 0.01	–	–
	Phe-11	1.5 ± 0.2	0.01 ± 0.03	–	–
	Glu-12	2.2 ± 0.2	0.7 ± 0.2	1.1 ± 0.2 (33%)	1.0 ± 0.2 (34%)
	Ser-13	2.1 ± 0.2	0.1 ± 0.1	1.0 ± 0.2 (33%)	–
	Leu-14	1.5 ± 0.2	0.01 ± 0.01	–	–
	Lys-15	1.9 ± 0.2	0.3 ± 0.2	1.3 ± 0.5 (50%)	1.1 ± 0.3 (34%)
	Ser-16	2.3 ± 0.1	0.4 ± 0.1	1.0 ± 0.1 (7%)	–
	Phe-17	1.8 ± 0.2	0.1 ± 0.1	–	–
httNTQ20_α–ins_2	htt^{NT}Q$_{20}$	2.3 ± 0.1	16 ± 1	14 ± 3 (100%)	2 ± 1 (80%)
	htt^{NT}	1.9 ± 0.2	4.6 ± 0.7	3 ± 2 (93%)	2 ± 1 (80%)
	Q$_{20}$	2.5 ± 0.2	12 ± 1	11 ± 3 (100%)	–
	Met-1	1.7 ± 0.3	0.05 ± 0.09	–	–
	Ala-2	2.1 ± 0.2	0.2 ± 0.1	–	–
	Thr-3	1.6 ± 0.2	0.03 ± 0.05	1.0 ± 0.1 (14%)	–
	Leu-4	1.6 ± 0.2	0.02 ± 0.05	–	–
	Glu-5	2.3 ± 0.2	0.9 ± 0.2	1.1 ± 0.2 (12%)	1.0 ± 0.1 (13%)
	Lys-6	2.1 ± 0.2	0.5 ± 0.2	1.3 ± 0.5 (54%)	1.2 ± 0.4 (43%)
	Leu-7	1.5 ± 0.2	0.01 ± 0.01	–	–
	Met-8	2.0 ± 0.2	0.3 ± 0.2	–	–
	Lys-9	2.4 ± 0.2	1.0 ± 0.2	1.0 ± 0.1 (10%)	1.0 ± 0.1 (9%)
	Ala-10	1.7 ± 0.2	0.01 ± 0.02	–	–
	Phe-11	1.7 ± 0.2	0.04 ± 0.06	–	–
	Glu-12	2.4 ± 0.2	0.8 ± 0.2	1.0 ± 0.2 (34%)	1.0 ± 0.1 (35%)
	Ser-13	2.1 ± 0.2	0.1 ± 0.1	1.0 ± 0.2 (45%)	–
	Leu-14	1.5 ± 0.2	0.01 ± 0.01	–	–
	Lys-15	2.0 ± 0.2	0.4 ± 0.2	1.5 ± 0.7 (49%)	1.2 ± 0.4 (38%)
	Ser-16	2.3 ± 0.2	0.2 ± 0.1	1.0 ± 0.1 (8%)	–
	Phe-17	1.7 ± 0.2	0.01 ± 0.01	–	–

Same caption as Table 11.IV for htt^{NT}Q$_{20}$ of httNTQ20_α–ins_1 and httNTQ20_α–ins_2.

11.5 Discussion

The huntingtin protein is related to Huntington's disease through pathways that could involve membrane interactions. Many experimental observations indicate that the huntingtin 17-amino-acid N-terminal (htt^{NT}), positioned just before the polyglutamine segment (Q$_N$), plays a crucial role in the modulation of its interactions with membrane [16, 398]. However, the dynamics and equilibrium properties of its interactions with a phospholipid bilayer as well as the effect of

Table 11.VI – Per residue htt^{NT}Q20–membrane interactions of httNTQ20_α-coil–ins.

Simulations	Amino acids	Insertion depth (nm)	SASA (nm^2)	H-bonds (no.)	Salt-bridges (no.)
httNTQ20_α-coil–ins	htt^{NT}Q$_{20}$	2.4 ± 0.2	16 ± 2	11 ± 3 (100%)	2±1 (88%)
	htt^{NT}	1.9 ± 0.2	4.6 ± 0.9	3 ± 2 (99%)	2 ± 1 (88%)
	Q$_{20}$	2.7 ± 0.2	12 ± 1	8 ± 3 (100%)	–
	Met-1	1.7 ± 0.2	0.03 ± 0.08	–	–
	Ala-2	2.2 ± 0.2	0.3 ± 0.1	–	–
	Thr-3	1.6 ± 0.2	0.02 ± 0.03	1.0 ± 0.1 (78%)	–
	Leu-4	1.6 ± 0.2	0.01 ± 0.03	–	–
	Glu-5	2.3 ± 0.2	0.8 ± 0.2	1.1 ± 0.2 (21%)	1.0 ± 0.1 (22%)
	Lys-6	2.1 ± 0.2	0.6 ± 0.2	1.4 ± 0.5 (60%)	1.3 ± 0.4 (53%)
	Leu-7	1.4 ± 0.2	0.1 ± 0.01	–	–
	Met-8	1.9 ± 0.2	0.2 ± 0.2	–	–
	Lys-9	2.4 ± 0.2	0.9 ± 0.2	1.0 ± 0.1 (18%)	1.0 ± 0.1 (17%)
	Ala-10	1.7 ± 0.2	0.01 ± 0.01	–	–
	Phe-11	1.6 ± 0.2	0.02 ± 0.4	–	–
	Glu-12	2.4± 0.2	0.7 ± 0.2	1.1 ± 0.3 (39%)	1.1 ± 0.2 (41%)
	Ser-13	2.0 ± 0.3	0.1 ± 0.1	1.0 ± 0.1 (37%)	–
	Leu-14	1.6 ± 0.2	0.01 ± 0.01	–	–
	Lys-15	2.1 ± 0.2	0.4 ± 0.3	1.1 ± 0.4 (47%)	1.0 ± 0.2 (39%)
	Ser-16	2.3 ± 0.3	0.3 ± 0.2	1.0 ± 0.1 (11%)	–
	Phe-17	1.8 ± 0.3	0.1 ± 0.1	–	–

Same caption as Table 11.IV for htt^{NT}Q$_{20}$ of httNTQ20_α-coil–ins.

the polyglutamine segment remain mostly unexplained at the atomic level. To our knowledge, no computational study of any sort was designed to look at these aspects to date. The main goal of this study is to fill this gap by investigating the atomic mechanisms responsible for the interactions between the htt^{NT}Q$_N$ peptide and a phospholipid bilayer using molecular dynamics simulations (Table 11.I).

11.5.1 Membrane interactions of htt^{NT}

Interactions of htt^{NT} with phospholipid membranes can lead to perturbations of their physical properties [71, 232, 319]. Circular dichroism (CD) experiments show that htt^{NT} populates α-helical configurations in hydrophobic environment such as created by the addition of detergents, TFE or unilamellar vesicles in aqueous buffer [16, 319, 459]. It is believed that, upon binding to the membrane, htt^{NT} becomes α-helical in order to partition its non-polar amino acids inside the hydrophobic core of the membrane. The resulting structure is expected to be a single α-helix with a well-defined amphipathic plane where the charged amino acids are all placed on the opposite side of the non-polar amino acids. This hypothesis however lacked atomistic models from either experimental or computational results until now.

In support of this hypothesis, our simulations show indeed that htt^{NT} interacts with the phos-

pholipid bilayer and can partition its non-polar amino acids inside the hydrophobic core of the membrane (Figure 11.1 and fourth column of Table 11.II). Our simulations also show that htt^{NT} can populate α-helical configurations while being adsorbed on or inserted in the membrane in agreement with results from CD experiments [16, 319, 459] (Figure 11.2 and fifth column of Table 11.II). While trajectories httNT_α_2 and httNT_α_3 sample largely random coil–turn configurations, other simulations starting from an initially inserted htt^{NT} (httNT_α–ins_1 and httNT_α–ins_2) lead us to conclude that these are structural intermediates in-route toward the stable α-helical state shown in Figure 11.4. When compared against our results on the stability of the α-helix in solution (Figure IV.3), we observe that the presence of the membrane significantly increases the stability of α-helical configurations for htt^{NT} ($16 \pm 28\%$ in solution vs. $80 \pm 17\%$ in membrane on average), in good agreement with a recent experiment that shows that the α-helical population goes from $\sim$10% to $\sim$80% upon stepwise addition of POPC/POPS (75 :25 molar ratio) or POPE/POPG (75 :25 molar ratio) vesicles in an aqueous buffer [319].

Furthermore, our structural results can be compared with the first nuclear magnetic resonance (NMR) atomistic model of htt^{NT} in apolar solution that has been obtained very recently [318]. Showing small variations amongst each other, their twenty atomic models have been obtained by solution NMR in a DPC micellar solution with their orientation discriminated on the basis of solid-state NMR constraints obtained for htt^{NT} on a POPC bilayer. In the following, we compare our results to their model #3, which satisfies best these latter constraints. In their model, htt^{NT} is unstructured from Met-1 to Glu-5 and is an α-helix from Lys-6 to Phe-17. We note however an artifact that could originate from the use of a micellar solution : residues 1 to 3 of the positively charged N-terminal, when oriented as measured experimentally on the phospholipid bilayer, are counterintuitively found directly in the hydrophobic core of the membrane in their model. Instead, our simulations of an initially inserted htt^{NT} on a phospholipid bilayer show, for their part, that the α-helix extends from Ala-2 to Ser-16 (Figure 11.5) with the charged N-terminal amine group staying outside the hydrophobic core of the membrane as expected (Figure 11.4). In terms of the peptide orientation with respect to the bilayer surface, our structural ensemble populates tilt angles that are near from their measured value of $103 \pm 5^\circ$ and rotational pitch angles that are smaller than their measured values of $137 \pm 5^\circ$ as shown in Table 11.III. It is tempting to relate this difference to the lipid composition of our membranes (POPE for us vs. POPC for them), however other experimental results obtained by the same

group suggest that the phospholipid composition does not influence much the orientation of the peptide on the membrane [319]. The discrepancy in the rotational pitch angle could be due to the fact that it is not clear if the orientation obtained by solid-state NMR is for the htt^{NT} monomer or for the dimer as it was suggested in this study that dimerization could occur through the electrostatic surface of the peptide forming an anti-parallel dimer [318]. The charged amino acids would then have a preferred orientation toward the other peptide resulting in a larger rotational pitch angle as observed in the NMR model, as opposed to our results showing that these residues are essentially more oriented toward the solvent with a pitch angle nearer from 90°.

Our results further complement the NMR model by providing a quantitative analysis of the peptide–membrane interactions at the atomic level in terms of insertion depth, surface accessible surface area, hydrogen bonds and salt-bridges (Tables 11.IV, 11.V and 11.VI). More specifically, this analysis reveals key residues stabilizing the htt^{NT} on the membrane : Glu-5, Glu-12, Lys-6 and Lys-15 are involved in salt-bridges, Thr-3 and Ser-13 are involved in hydrogen bonds, and Leu-4, Leu-7, Phe-11 and Leu-14 are essentially buried. All these residues have been previously recognized to be crucial for the localization, aggregation and interaction partners of huntingtin N-terminal. For instance, phosphorylation or phosphomimetic mutations of Ser-13 increases its nuclear localization as well as its degradation by the proteasome and lysosome [175, 463], phosphorylation of Thr-3 affect its aggregation reducing the neurotoxicity [4], SUMOylation of the lysines reduces its ability to form visible SDS-insoluble aggregates [443], mutations of Leu-4 and Phe-11, which are part of the nuclear export sequence of huntingtin N-terminal along with Leu-7, Leu-14 and Ser-16 [537], to alanines abrogate its ability to target the endoplasmic reticulum [16], and mutations of Glu-5 and Glu-12 to alanines increase vesicle targeting [16]. Here, our results suggest that mutations and post-transcriptional modifications of these residues, involved in interactions with the membrane, are likely to affect the structure and orientation of the htt^{NT} peptide on the membrane.

As well as providing information on the structural properties of htt^{NT}, our simulations also reveal its insertion dynamics inside a phospholipid bilayer. We observe that it follows, similarly to α-helical membrane-active peptides [30, 33, 385], four main steps : an approach step driven by the charged amino acids of the peptide, a reorganization step during which htt^{NT} undergoes structural changes leading to an anchoring step that can be initiated by either phenalylalanines and an insertion step during which the other non-polar amino acids are partitioned inside the hy-

drophobic core of the membrane as reorganization continues (Figure 11.1). The details of these steps can vary from one trajectory to the other as to the precise sequence of the residues interacting first with the membrane and of the non-polar residues partitioning (Table 11.II). We also observe that some of the trajectories get trapped for hundreds of nanoseconds between the third and fourth steps waiting for specific conformational events that are necessary for the partitioning of more non-polar amino acids as well as the formation of the stable α-helix conformation shown in Figure 11.4. This bottleneck is due to strong electrostatic interactions between the charged amino acids of the peptide and the phospholipids' head group that must be temporarily broken in order for the peptide to proceed toward a more structured and inserted state. As a result, the peptide–membrane Coulomb energy fluctuates considerably as the peptide proceeds towards this state (Figure IV.6).

11.5.2 Impact of the glutamines

The number of glutamines in the Q_N region is an important modulator of aggregation and fibrillation in solution as shown from both experiments [90, 91] and simulations [23, 306, 485]. This region also seems to have a role for the membrane interactions of huntingtin [71, 230, 232]. However, to our knowledge, the atomistic details of the effects of this region on the structure and dynamics of the interactions between $htt^{NT}Q_N$ and a membrane have not been unveiled by experiment or simulation yet.

Our simulations on $htt^{NT}Q_{20}$ show that its insertion dynamics is described by the same four main steps as without the Q_N region : approach, reorganization, anchoring and insertion (Figure 11.1). As for htt^{NT}, the details of each step can vary from one simulation to the other (Table 11.II). For instance, the Q_N region is within the first part of the peptide to contact the membrane as can the charged amino acids of the htt^{NT} region. Once adsorbed, the glutamines' side-chain form hydrogen bonds with the phospholipids' head group : 5 ± 2 in httNTQ20_α_1 and 10 ± 3 in httNTQ20_α_2 averaged over the last 200 ns. These interactions between the Q_N region and the membrane provide a stable scaffold for the partitioning of the non-polar amino acids inside the hydrophobic core of the membrane. As a result, the final configurations of $htt^{NT}Q_{20}$ are more inserted (Figure 11.3) and structured (Figure 11.2) than htt^{NT}.

Interestingly, the trajectories of htt^{NT} and $htt^{NT}Q_{20}$ starting from an initially inserted α-helical peptide display striking similarities (Figure 11.4 vs. Figure 11.7). For instance, the

htt^{NT} region stays well-structured in a single α-helix starting from Ala^2 (Figure 11.5). Also, the position densities of the non-polar residues inside the hydrophobic core of the membrane are similar (Figure 11.6). The orientation of the htt^{NT} region is also largely preserved, with a small shift towards smaller tilt and pitch angles (Table 11.III). We observe that these shared features originate from comparable peptide–membrane interactions as seen by comparing Table 11.IV to Table 11.V and are mostly independent of the conformation adopted by the Q_N region (Table 11.VI). Overall, these results suggest that the Q_N region does not significantly modulate the final structure of the htt^{NT} region. Its main role is rather to modulate the insertion dynamics by securing the peptide on the membrane surface through hydrogen bonds. Moreover, as the amyloidogenic Q_N region stays on the surface of the membrane and as it does not interact with the htt^{NT} region (Figure 11.7 and Figure IV.11, and Tables 11.V and 11.VI), another important role for it could occur during oligomerization by forming extended β-sheet structures with the Q_N region of other htt$^{NT}Q_N$ peptides thus promoting fibrillation. The Q_N regions could be brought in close contact by simple two-dimensional diffusion of the anchored htt^{NT} region on the membrane. The dimer complex could be then further stabilized by the formation of salt-bridges between the charged amino acids of the two htt^{NT} regions [318]. It can then be expected that the membrane, as for other amyloid proteins [263], could enhance fibrillation of the htt$^{NT}Q_N$ peptide.

Experimental results show that huntingtin and htt$^{NT}Q_N$ interact more strongly with phospholipid membrane as the number of glutamines in Q_N increases [71, 230, 232]. To investigate this trend, we simulate an intermediate length for the Q_N region : htt$^{NT}Q_{10}$ (simulations httNTQ10_α_1, httNTQ10_α_2 and httNTQ10_coil). Unfortunately, it becomes rapidly obvious that few clear trends can be extracted between the observations on htt^{NT}, htt$^{NT}Q_{10}$ and htt$^{NT}Q_{20}$ as each system samples very different trajectories on their own (Table 11.II and Figures 11.1, IV.5, IV.9 and IV.12). We do note however that the approach step of htt$^{NT}Q_{10}$ is very similar to that of htt$^{NT}Q_{20}$ since its glutamines are first to make contact with the membrane. Its trajectories are trapped in the second step due to strong interactions of the glutamines and the charged amino acids with the phospholipids' head group (Figure IV.12) as for some trajectories of htt$^{NT}Q_{20}$ (Figure IV.9). Finally, our results show that the averaged number of hydrogen bonds between the peptide's side-chains and the membrane for the systems starting from a single α-helix in solution increases from htt^{NT} (3 ± 4 on average) to htt$^{NT}Q_{10}$ (8 ± 3 on average) and

to $\text{htt}^{\text{NT}}\text{Q}_{20}$ (12 ± 7 on average). The trajectories of an initially inserted peptide also show this trend : from 3 ± 1 for htt^{NT} to 14 ± 4 for $\text{htt}^{\text{NT}}\text{Q}_{20}$ (on average). The Q_{N}-length dependence could then originate in part from the increase of the number of hydrogen bonds between the glutamines' side-chain and the phospholipids' head group as the Q_{N} region gets longer securing more efficiently the N-terminal on the membrane surface.

11.6 Conclusion

The $\text{htt}^{\text{NT}}\text{Q}_{\text{N}}$ region at the N-terminal of huntingtin is thought to be crucial for huntingtin physiological and pathological roles by modulating its localization to membrane-containing organelles in the cell. The molecular dynamics simulations presented here unveil the structures and dynamics of the interactions between $\text{htt}^{\text{NT}}\text{Q}_{N}$ and phospholipid membranes at the atomic level. Its insertion dynamics on a membrane is typical of α-helical membrane-active peptides as it follows four main steps – approach, reorganization, anchoring and insertion – that are very diverse at the atomic level, and it forms a stable α-helix essentially parallel to the surface of the membrane. More specifically to the htt^{NT} peptide, its structure and orientation are not significantly modulated by the presence of the Q_{N} region whose effects are observed elsewhere. First, it provides a stable scaffold, through hydrogen bonds with the phospholipids' head group, for the partitioning of the non-polar amino acids of the htt^{NT} region inside the hydrophobic core of the membrane. Second, the Q_{N} region, being able to form amyloid fibrils in solution, could also promote oligomerization and fibrillation on the membrane by recruiting other $\text{htt}^{\text{NT}}\text{Q}_{\text{N}}$ as our simulations show that it stays accessible to the solvent above the phospholipids' head group. The dimer could then be further stabilized by the formation of salt-bridges between the htt^{NT} regions. Finally, our results complement previous experiments by providing a quantitative analysis of the relationship between the $\text{htt}^{\text{NT}}\text{Q}_{\text{N}}$ peptide and the phospholipid membrane by combining insertion depth, solvent accessible surface area, hydrogen bond and salt-bridge measurements. This analysis reveals the key residues of the htt^{NT} peptide for salt-bridges (Glu-5, Glu-12, Lys-6 and Lys-15), hydrogen bonds (Thr-3 and Ser-13) and non-polar (Leu-4, Leu-7, Phe-11 and Leu-14) contributions to its stability on the membrane as well as those that are less crucial by staying essentially accessible to the solvent with marginal interactions with the membrane (Lys-9 and Ser-16).

11.7 Supporting Material

Additional Supporting Information may be found in Annexe IV.

11.8 Author contributions

S.C., G.W. and N.M. designed the research ; S.C. performed the research ; S.C. contributed analytic tools ; S.C., G.W. and N.M. analyzed the data ; and S.C., G.W. and N.M. wrote the manuscript.

11.9 Acknowledgments

This work was funded in part by the Canada Research Chairs program, the Fonds Québécois de Recherche sur la Nature et les Technologies (FQRNT), the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Fonds de Recherche en Santé du Québec (FRSQ). The simulations were executed on the supercomputers of Calcul Québec. The authors thank Evgeniy S. Salnikov and Burkhard Bechinger for helpful discussions.

CHAPITRE 12

INTERACTIONS MEMBRANAIRES DU N-TERMINAL DE HUNTINGTINE – PARTIE 2

Probing the Huntingtin 1-17 Membrane Anchor on a Phospholipid Bilayer by Using All-atom Simulations

Sébastien Côté[†], Vincent Binette[†], Evgeniy S. Salnikov[‡], Burkhard Bechinger[‡], and Normand Mousseau[†]

† Département de Physique and Groupe d'étude des protéines membranaires (GÉPROM), Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec), Canada

‡ Université de Strasbourg/Center National de la Recherche Scientifique, UMR7177, Institut de Chimie, Strasbourg, France

Article paru : Côté, S., V. Binette, E. S. Salnikov, B. Bechinger, and N. Mousseau. 2015. Probing the huntingtin 1-17 membrane anchor on a phospholipid bilayer by using all-atom simulations. *Biophys. J.* 108 :1187-1198.

12.1 Abstract

Mislocalization and aggregation of the huntingtin protein are related to Huntington's disease. Its first exon – more specifically the first 17 amino acids (Htt17) – is crucial for the physiological and pathological functions of huntingtin. It regulates huntingtin's activity through posttranslational modifications and serves as an anchor to membrane containing organelles of the cell. Recently, the structure and orientation of the Htt17 membrane anchor were determined using a combined solution and solid-state NMR approach. This prompted us to refine this model by investigating the dynamics and thermodynamics of this membrane anchor on a POPC bilayer using all-atom, explicit solvent molecular dynamics and Hamiltonian replica exchange. Our simulations are combined with various experimental measurements to generate a high-resolution atomistic model for the huntingtin Htt17 membrane anchor on a POPC bilayer. More precisely, we observe that the single α-helix structure is more stable in the phospholipid membrane than the NMR model obtained in the presence of DPC detergent micelles. The resulting Htt17 monomer has its hydrophobic plane oriented parallel to the bilayer surface. Our results further

unveil the key residues interacting with the membrane in terms of hydrogen bonds, salt-bridges and non-polar contributions. We also observe that Htt17 equilibrates at a well-defined insertion depth and that it perturbs the physical properties – order parameter, thickness and area per lipid – of the bilayer in a manner that could favor its dimerization. Overall, our observations reinforce and refine the NMR measurements on the Htt17 membrane anchor segment of huntingtin that is of fundamental importance to its biological functions.

12.2 Introduction

Huntingtin is an ubiquitous protein of 3114 amino acids fundamental to the embryonic development [143] and is involved, for example, in intracellular vesicular and organelles trafficking [79] such as regulating autophagy in response to endoplasmic reticulum stress [15], protein scaffolding [231], transcription and axonal transport [535]. It has attracted considerable attention over the last decades due to its relation to Huntington's disease, a CAG/glutamine repeat disorder [26, 168, 358, 539]. Indeed, it contains a segment of consecutive glutamines (Q_N) at its amino-terminus that leads, when expressed with more than approximately 36 repeating residues, to the aggregation of huntingtin causing, in a dominant manner, deleterious effects through various nuclear and extranuclear pathways [38, 395]. It has been observed that amino-terminus fragments generated by proteolytic cleavage of huntingtin accumulate in the nucleus and are involved in the pathogenesis of Huntington's disease [134, 242, 392]. The first exon of huntingtin is sufficient to reproduce neurological phenotypes and aggregation features characteristic of Huntington's disease *in vivo* [120, 304] and *in vitro* [99, 350, 526]. It is composed of three main regions : an amphipathic segment of 17 amino acids (Htt17), a segment of consecutive glutamines (Q_N), and a segment rich in prolines.

The Htt17 segment of the first exon is of fundamental importance. It acts as a membrane anchor that modulates the localization of huntingtin to specific membrane-containing organelles of the cell such as the endoplasmic reticulum [16], the mitochondria and the Golgi [398]. Htt17 also acts as a nuclear export sequence (NES) that regulates the localization of huntingtin between the cytoplasm and the nucleus [302, 536]. It interacts with the TRiC chaperonin to suppress the aggregation of huntingtin [459]. Moreover, it can undergo posttranslational modifications such as SUMOylation [443] and phosphorylation [4, 14, 175, 463] that modulate the toxicity, clearance, localization and function of huntingtin. This small segment also accelerates the fibrilla-

tion kinetics [275, 459] and changes the nucleation and the oligomer structures [209, 210] of the amyloidogenic Q_N segment [501]. These observations indicate that Htt17 is crucial to huntingtin physiological and pathological functions.

In aqueous solution, circular dichroism (CD) experiments show that the Htt17 segment populates 10 to 50% of α-helix depending on the aqueous buffer used [16, 319, 459, 504]. Solution nuclear magnetic resonance (NMR) studies further indicate that it does not sample any stable secondary structure motif [459]. Taken together, these results suggest that Htt17 likely populates a wide range of helix/coil configurations. In contrast, Htt17 appears to form an α-helix in a crystal structure of a fusion protein consisting of the maltose-binding protein and huntingtin exon 1 (MBP-Htt17Q_{17}-ex1) [239]. Complementing these *in vitro* experiments, various computational methods investigated the configurational ensemble of Htt17 and its impact on Q_N at the atomic level. While some observe that Htt17 forms an α-helix in aqueous solution by its own [233] and attached to Q_N [137], most observe that Htt17 is for the most part unstructured by its own [110, 111, 401, 504] and attached to Q_N [260, 335, 504]. Htt17 also modulates the structure [110, 504] and the aggregation [403] of Q_N oligomers.

In apolar solution, there is a significant increase of the α-helix population as shown by CD spectroscopy when adding TFE, DPC detergent micelles, or lipid vesicles to aqueous buffers [16, 319, 459]. Early studies postulated that Htt17's ability to form an α-helix upon binding to a phospholipid bilayer is encoded in its amphipathic amino acid sequence [16]. Indeed, as an α-helix, the Htt17 has all its non-polar amino acids lying on one side separated from the charged amino acids by a well-defined hydrophobic plane. Recent solution NMR experiments by Michalek and co-workers further unveiled the atomic structure of Htt17 in a DPC micelles solution, showing that it forms a well-structured α-helix from residues 6 to 17, while the first five residues are disordered [319]. Following these results, the orientation of this model on POPC bilayers has been determined using the ^{15}N chemical shifts of Leu7, Phe11 and Phe17 as well as the deuterium quadrupolar splitting of $^{2}H_3$-Ala10 obtained by solid-state NMR [318]. The Htt17 insertion depth on a phospholipid bilayer has also been recently investigated from the fluorescence intensity quenching of three key residues (Met1, Phe11 and Phe17) mutated to tryptophan [317]. Other experiments focused on the membrane interactions of Htt17 when linked to Q_N revealing, for instance, that Htt17 is crucial for Q_N binding and aggregation on membrane bilayers [72], and that it binds more favorably to curved [82] and acidic phospholipid-containing [230] bilayers

as the Q_N segment's length increases [71, 230].

The structure and dynamics of Htt17 on a phospholipid membrane at the atomic level has also been investigated using atomistic molecular dynamics (MD) simulations. The binding of Htt17, Htt17Q_{10} and Htt17Q_{20} as well as their stability as an α-helix on a POPE membrane has been simulated [111]. Upon binding to the membrane, the Htt17 segment partitions its non-polar amino acids inside the hydrophobic core of the membrane, while Q_N remains above the phospholipid head groups. The role of the Q_N region is to stabilize Htt17 membrane anchor easing its insertion as a stable single α-helix in agreement with previous experimental observations [318, 319]. Another group focused on the binding of KKQ_{35}KK, KK$Q_{35}P_{11}$KK, Htt17Q_{35}KK, Htt17$Q_{35}P_{11}$KK from disordered states on a DOPC bilayer [335]. They observed that binding of these sequences leads to perturbations of the bilayer physical properties. Moreover, Htt17 enhances the membrane interactions of huntingtin N-terminal in agreement with previous experiments [72]. Both of these computational studies show that Htt17 folding on a bilayer from disordered states occurs beyond the microsecond timescale and involves a slow configurational change [111, 335].

With the publication of the NMR model [318, 319], we decided to refine the atomistic picture of Htt17 interactions and dynamics on a POPC bilayer using all-atom molecular dynamics (MD) and Hamiltonian replica-exchange (HREX) simulations. We specifically analyze questions of fundamental importance that still need to be addressed : (i) Is the NMR model obtained in the presence of DPC micelles more structured on a POPC bilayer ? (ii) What are the most significant Htt17–membrane interactions ? (iii) What is the insertion depth of each residue ? and (iv) What are the bilayer perturbations induced by a structured Htt17 membrane anchor ? Such knowledge is valuable to understand the atomistic basis of huntingtin interaction with membrane-containing organelles of the cell through its amino-terminus membrane anchor (Htt17). It further provides the necessary information to develop rationals for modulating the localization and aggregation of huntingtin in the cell at the atomic level.

12.3 Material and Methods

In this study, we combine molecular dynamics (MD) and Hamiltonian replica exchange (HREX) simulations respectively of 26 μs and 20 μs in total, as summarized in Table 12.I, to investigate the configurational ensemble of the 17-amino-acid amino-terminus segment (Htt17) of

the huntingtin protein on a 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) bilayer. The amino acids sequence of Htt17 is : MATLEKLMKAFESLKSF-NH_2, where we use an amidated carboxy-terminus as in the NMR experiments to which we compare and combine our results [318, 319].

All simulations are performed using the Gromacs software (version 4.6.5) [42, 185, 387, 478] combined with the PLUMED plug-in (version 2.0.2) for the HREX simulations [74, 467]. The AMBER99sb*-ILDN protein forcefield [382] is combined with the Berger phospholipid forcefield [102] as justified in the Supporting Material. The HREX algorithm enhances the sampling by executing simultaneous simulations called replicas with different Hamiltonian (energy) that can exchange, at a given frequency, as specified by the REST2 protocol for which only protein/protein and protein/solvent interactions are scaled [492]. In our case, the solvent is the water molecules and the phospholipids, therefore water/water, water/phospholipids and phospholipids/phospholipids interactions are not scaled. This protocol has been tested on trpcage and a β-hairpin showing a significantly lower computational cost and better sampling than temperature replica exchange [492]. Moreover, contrary to temperature replica exchange, this method enhances the sampling of the conformational space at room temperature, which is a necessary condition to conserve the integrity as well as the physical properties of the phospholipid bilayer. For our HREX simulation, we use 16 scales generated by a geometric distribution (and numbered from 1 to 16 in the following order) : 1.0, 0.92, 0.85, 0.79, 0.73, 0.67, 0.62, 0.57, 0.53, 0.49, 0.45, 0.41, 0.38, 0.35, 0.33 and 0.3. An exchange between each neighboring scale is attempted every 2 ps resulting in an exchange rate of 10–30%. More details on this method and on our simulation parameters are provided in the Supporting Material.

The simulations performed in this paper are summarized in Table 12.I. We use two initial states : the NMR model obtained in a the presence of DPC micelles (PDB : 2LD2) [319], and a single α-helix. The NMR model that we use is the third structure in the 2LD2 PDB as it best satisfies the chemical shift constraints obtained by solid-state NMR on a POPC bilayer [318]. The other structures are very similar to the third one as STRIDE [164] shows that they all populate disordered configurations from residues 1 to 5 and an α-helix from residue 6 to 16 [319]. For its part, the single α-helix, as its name indicates, has a single α-helix running from residue 2 to 16. We use the InflateGRO2 procedure with the standard parameters [413] to remove all atomic clashes of the inserted Htt17. Initially, the center-of-mass distance between the backbone atoms

of residues 6 to 17 and the bilayer is 1.8 nm, and the Htt17 lies parallel to the surface of the membrane with its non-polar residues facing the hydrophobic core of the membrane. We also tested two additional initial positions at 2.0 and 2.5 nm using HREX simulations only. Each MD trajectory and HREX replica starts from a different velocity distribution following an energy minimization and a NPT equilibration of 10 ns.

12.4 Results

In this section, we analyze more specifically three simulation sets : 11 MD simulations starting from the NMR model (Htt17_nmr) as well as 11 MD and a HREX simulation starting from a single α-helix (Htt17_α and Htt17_α_hrex, respectively) as summarized in Table 12.I. We compare, discuss and combine our results with the experimental measurements obtained by solid-state and two dimensional solution NMR [318, 319] as well as fluorescence [317] spectroscopy. Further comparisons with experimental studies are postponed to the Discussion section.

Table 12.I – Summary of the performed simulations.

Simulations	Phospholipid forcefield	Type	Initial configuration	Time (ns)
Htt17_nmr	Berger	MD	NMR	1000 ns $\times$ 11
Htt17_α	Berger	MD	α	1000 ns $\times$ 11
Htt17_nmr_slipids	SLIPIDS	MD	NMR	1000 ns $\times$ 2
Htt17_α_slipids	SLIPIDS	MD	α	1000 ns $\times$ 2
Htt17_nmr_hrex	Berger	HREX	NMR	500 ns $\times$ 16 $\times$ 1
Htt17_α_hrex	Berger	HREX	α	250 ns $\times$ 16 $\times$ 3

All simulations are done in the NPT ensemble and are started from the same hexagonal prism periodic cell ($\alpha = 90°$, $\beta = 90°$, $\gamma = 120°$, a = 7.73 nm, b = 7.73 nm, c = 9.24 nm) containing 8982 water molecules and 167 phospholipids. For the HREX simulation, the number of replicas is 16 as specified in the last column. The AMBER99sb*-ILDN/Berger simulations are presented in the main text, and those using AMBER99sb*-ILDN/SLIPIDS in the Supporting Material. The Htt17_nmr_hrex and two of the Htt17_α_hrex simulations are presented in the Supporting Material.

12.4.1 Htt17 structural stability

We observe that Htt17 remains largely α-helical, particularly from residues 7 to 16 in the three simulation sets, with a probability greater than 50% as shown in Figure 12.1. The two initial

states evolve differently, however. While the NMR model (Htt17_nmr) loses helicity and shows significant fluctuation at residues 6 to 9, the single α-helix (Htt17_α and Htt17_α_hrex) is much better preserved overall with a per residue α-helical probability greater than 80% for residues 2 to 15 with only very small fluctuations. The efficiency of the sampling enhancing HREX simulation is supported by the fact that the replicas diffuse well (10–30% of exchange acceptance) and that the peptide unfolds at larger scales corresponding to reduced interactions (Figure V.1). Furthermore, starting from different insertion depths does not change the stability of the α-helix that converges to the same equilibrium properties such as secondary structure, water accessibility, insertion depth and membrane properties (Figure V.2).

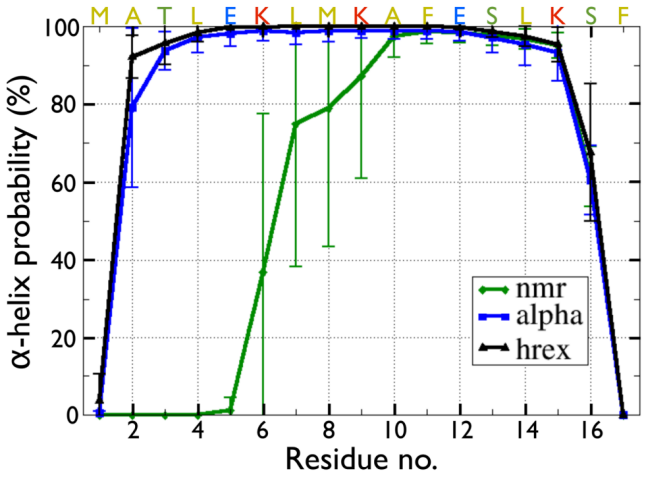


Figure 12.1 – Per residue α-helix probability of Htt17. Results for the MD simulations starting from a single α-helix and from the NMR model determined in the presence of DPC micelles are respectively shown by blue squares and green diamonds. The HREX simulation is shown by black triangles. Initially, according to STRIDE [164], the NMR model has an α-helix from residues 6 to 16 and the alpha model from residues 2 to 16. The HREX simulation is launched from the alpha model. The value and the error bar for each residue are respectively the average and the standard deviation over all the averages on the 250–1000 ns time interval obtained from 11 independent MD simulations (Table 12.I). For HREX, the average and standard deviation are computed at the unscaled replica on the 50–250 ns time interval using 20-ns time windows. See Figures V.2 and V.3 for the results on the two additional HREX simulations starting from the single α-helix at different insertion depths and the one starting from NMR model, respectively.

By comparison with the single α-helix, the NMR model is less stable as indicated by the relatively large error bars at residues 6 to 9. This can be explained by the fact that the helical structure unfolds (turn and coil) for these residues in some of the MD trajectories. A HREX simulation starting from the NMR model (Htt17_nmr_hrex) shows even more unfolding of the

conformation initially (Figure V.3). Interestingly, however, the Htt17 refolds in this simulation towards a single α-helix near the end of the simulation. The sampling enhancing technique HREX seems, in contrast to the MD simulations, to be efficient enough to refold the Htt17 peptide on bilayer. A longer simulation time, currently beyond our accessible computational resources, will be needed to reach convergence and confirm this observation.

These results suggest that Htt17 could be more structured on a bilayer than the micelle NMR model that has rather flexible amino-terminus residues [319]. This is supported by tryptophan fluorescence indicating a relatively deep penetration of the first residue of Htt17 [317]. Moreover, solid-state NMR measurements of Htt17 reconstituted into oriented phospholipid bilayers show a small non-zero ^{2}H splitting for Ala2, suggesting structure/interaction with the membrane (E. S. Salnikov & B. Bechinger, unpublished observation). This small splitting could either be due a high degree of motion and/or an alignment of the C_α–C_β vector close to the magic angle. Our simulations also find, in agreement with ssNMR, a small quadrupolar splitting for Ala2 (Table 12.II). This confirms that the magnitude of this value is due in part to a stable α-helical conformation around this residue, and in part to a higher degree of motions for Ala2 compared to residues 6-15 as indicated by the fluctuations in the α-helical content (Figure 12.1).

Table 12.II – Orientation of Htt17 on the bilayer.

	$^2H_3$-Ala^2 kHz	^{15}N-Leu^7 ppm	$^2H_3$-Ala^{10} kHz	^{15}N-Phe^{11} ppm	^{15}N-Leu^{14} ppm	^{15}N-Phe^{17} ppm	Tilt $^\circ$	Pitch $^\circ$
Htt17_nmr	14 ± 20	110 ± 28	26 ± 12	91 ± 3	77 ± 6	80 ± 15	94 ± 5	76 ± 5
Htt17_α	-7 ± 10	71 ± 2	30 ± 6	85 ± 1	66 ± 1	87 ± 4	87 ± 5	85 ± 5
Htt17_α_hrex	-4 ± 7	68 ± 3	27 ± 8	84 ± 1	65 ± 2	92 ± 5	91 ± 5	95 ± 5
ssNMR (exp)	5 ± 2	71.2 ± 1.7	11 ± 2.5	78.9 ± 1.5	73.3 ± 1.2	88.2 ± 0.9	103 ± 5	137 ± 5

The ^{15}N chemical shifts of Leu7, Phe11, Leu14, and Phe17 as well as the $^2H_3$ quadrupolar splittings of Ala2 and Ala10 are used to extrapolate the orientation of the membrane anchor on the bilayer in terms of tilt and rotational pitch angles [5, 36, 408]. The ssNMR measurements of Htt17 were done on a POPC bilayer and the tilt and pitch angles were determined for the NMR structure of Htt17 in the presence of DPC micelles [318, 319]. For the MD simulations, values and their error bar are respectively the average and the standard deviation over all the averages on the 250–1000 ns time interval obtained from 11 independent simulations (Table V.I). For the HREX simulation, the average and standard deviation are computed at the unscaled replica on the 50–250 ns time interval using 20-ns time windows. See Table V.I for the results on the two additional HREX simulations starting from the single α-helix at different insertion depths.

As expected, the remaining part of the structure, which forms a stable α-helix, is very similar to the solution NMR model [318, 319] with a backbone RMSD on residues 6 to 16 of 0.11 $\pm$

0.07 nm (Htt17_nmr), 0.07 $\pm$ 0.01 nm (Htt17_α) and 0.07 $\pm$ 0.01 nm (Htt17_α_hrex). We also note that residue 17 could be considered part of the helix as main-chain hydrogen bonds between this residue and residues 13 or 14 are populated 68 $\pm$ 5% of the time in Htt17_nmr, 70 $\pm$ 4% in Htt17_α and 69 $\pm$ 7% in Htt17_α_hrex. This is in agreement with the NMR spectra that show NOEs up to residue 17 and the ssNMR spectra that exhibits an anisotropic ^{15}N chemical shift at this residue [318, 319].

12.4.2 Htt17 orientation

In terms of orientation, we compare our simulations to ssNMR measurements on the ^{15}N chemical shifts of Leu7, Phe11 Leu14 and Phe17, and the $^{2}H_3$ quadrupolar splittings of Ala2 and Ala10 of Htt17 on a POPC bilayer [318, 319]. These values depend respectively on the orientation of the C–N and N–H bonds (^{15}N), and the C_α–C_β bond (^{2}H) with respect to the bilayer normal that is aligned in the same direction as the magnetic field [5, 36, 408]. The computations are described in detail in the Supporting Material.

The averaged chemical shifts and quadrupolar splittings sampled during the Htt17_nmr, Htt17_α and Htt17_α_hrex simulations are presented in Table 12.II. The ^{15}N-Phe^{11}, ^{15}N-Leu^{14} and ^{15}N-Phe^{17} chemical shifts and the $^{2}H_3$-Ala^{2} quadrupolar splitting for the three simulation sets show good agreement with ssNMR measurements. The signal for ^{15}N-Leu^{7} favors rather the α-helical configuration over the NMR model that becomes disordered around this residue in our simulations (Htt17_nmr in Figure 12.1). Strangely, however, the $^{2}H_3$-Ala^{10} quadrupolar splitting sampled in the three simulation sets is significantly larger than the experimental measurement. Using a different forcefield (AMBER99sb*-ILDN/SLIPIDS) yields a value for the $^{2}H_3$-Ala^{10} quadrupolar splitting that is closer to the experiment, but that is still different (17 $\pm$ 1 kHz vs. 11 $\pm$ 1kHz for ssNMR, Table V.I). For the single α-helix, this difference could be due to very local structural fluctuations of modest amplitude as the chemical shifts of Ala10's neighboring residues (^{15}N-Leu^{7}, ^{15}N-Phe^{11} and ^{15}N-Phe^{17}) agree with the experimental values. Moreover, this part of the peptide samples similar secondary structure propensity in all simulations, suggesting a very stable global conformation (Figure 12.1). These observations are reinforced by two supplementary HREX simulations starting from a less inserted single α-helix that yield very similar chemical shifts and quadrupolar splitting (Table V.I).

We can further describe the orientation of Htt17 on a bilayer in terms of the tilt and rotational

pitch angles as shown in Table 12.II. These angles are defined as a rotation of Htt17 structure's amphipathic plane, initially in the z-x plane, around the negative z-axis (rotational pitch) followed by the rotation around the negative y-axis (tilt) as shown in Figure 12.2A. Using this definition, a rotation of 90° tilt and 90° pitch results in the hydrophobic plane being parallel to the membrane surface (xy-plane). These angles can be extracted by the values of ^{15}N chemical shifts and ^{2}H quadrupolar splitting. In particular, the tilt angle depends more strongly on the ^{15}N shift [36]. Here, the good agreement between simulations and ssNMR in the ^{15}N chemical shifts indicates that the helix axis of Htt17 has an in-plane orientation in the membrane (Table 12.II).

The rotational pitch angle, for its part, is a more difficult quantity to extract as it depends on the ^{2}H quadrupolar splitting, a value that is very sensitive upon local structural changes as observed from experiments and simulations [35, 359]. In our simulations, the hydrophobic plane of Htt17 is oriented parallel to the membrane surface (pitch $\sim 90°$) as expected for an amphipathic monomer such as Htt17 [5]. This value differs significantly from the experimentally determined pitch angle (137°) that suggests that the charged residues of Htt17 are preferentially oriented toward the right when looking along the helix axis through the carboxy-terminus (such a point of view is shown in Figure 12.2B). This asymmetrical orientation was justified by the possible formation of Htt17 dimers on the bilayer during the ssNMR experiment through electrostatic interactions [318, 319]. However, introducing small conformational changes to the micellar NMR structure could also result in a more symmetric alignment while at the same time satisfying all solid-state NMR constraints.

Our simulations can further be used to characterize the global motion of the Htt17 helix. We observe that the wagging motion along the helix long axis spreads by a standard deviation of about 8–10°, while the wobbling motion around the helix long axis spreads by a standard deviation of about 14° (data not shown). These angular deviations are quite close to those used for the restriction analysis of the solid-state NMR data where standard deviations of 10° and 18°, respectively, were used for the analysis [318, 319]. This motional regime used for the analysis of tilt and pitch angles, which is confirmed by the MD results, represents better the situation in a liquid crystalline bilayer than a completely static peptide, which led to differences of up to about 5°.

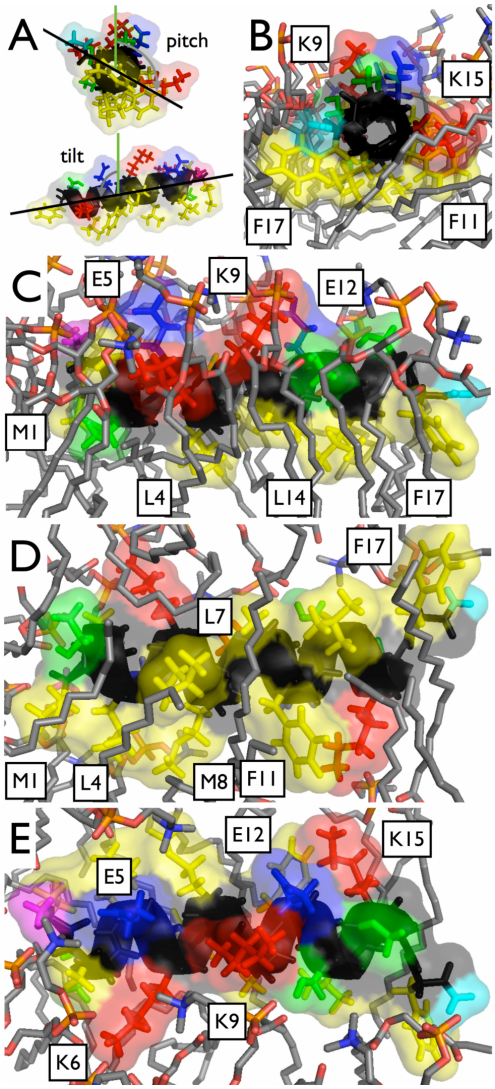


Figure 12.2 – Representative structure of the average orientation of Htt17 on a POPC bilayer from the MD trajectory Htt17_α. The corresponding ^{15}N chemical shifts would be 71 (Leu7), 85 (Phe11) and 87 (Phe17) ppm, and the $^{2}H_3$–Ala^{10} quadrupolar splitting is 30 kHz corresponding to a tilt angle of 87 $\pm$ 5° and rotational pitch angle of 85 $\pm$ 5°. (A) Graphical definition of the tilt and rotational pitch angles, (B) view from the carboxy-terminus, (C) view from the side, (D) view form the membrane, and (E) view from the aqueous solution. The non polar, negatively charged, positively charged and polar amino acids of Htt17 are respectively shown in yellow, blue, red and green. The backbone atoms are displayed in black, and the amino- and carboxy-terminus are pink and teal, respectively.

12.4.3 Htt17 interactions with the bilayer

After characterizing Htt17's structural and topological properties, we now examine its interactions with the POPC bilayer. In terms of solvent accessibility, we observe the sequestration of

key non-polar residues such as Leu7, Phe11, Leu14 and Phe17 for the three simulation sets (Figure 12.3). Moreover, residues involved in frequent hydrogen bonds or salt-bridges with the phosphate group of the phospholipids (Lys6, Ser13 and Lys15, see Table 12.III) are also more isolated from the solvent. As expected, those that weakly interact with the bilayer (Glu5, Lys9 and Glu12) are highly accessible to the solvent. For the Htt17_α and Htt17_α_hrex simulations, residues 1 to 4 are less accessible to the solvent compared to Htt17_nmr for which these residues are and stay disordered (Figure 12.1). Measurements of Htt17 insertion depth in the bilayer support these observations (Figure 12.3).

Our simulations can be combined with recent experimental observations [317, 318]. First, the addition of hydrophilic (Mn^{2+} ions) or hydrophobic (16-doxyl-stearic acid) paramagnetic relaxation reagents in the presence of DPC micelles indicate that residues Leu7, Met8, Ala10 and Phe11 of Htt17 are oriented toward the hydrophobic environment of the micelle [318]. Our results correlate well with these measurements and confirm that these residues are also strongly involved in the association of Htt17 with a POPC bilayer (Figure 12.3). Second, using fluorescence intensity quenching, the insertion depth of residues Met1, Phe11 and Phe17 mutated to tryptophan (Trp) was measured in the presence of POPC–POPS vesicles containing brominated POPC lipids [317]. The relative positions of M1W, F11W and F17W in Htt17 were respectively measured to 1.12 $\pm$ 0.14 nm, 0.73 $\pm$ 0.12 nm and 1.08 $\pm$ 0.12 nm from the center of the bilayer. Our results from the Htt17_α and Htt17_α_hrex simulations agree with the measurement on Met1, while Phe17 and Phe11 are slightly farther from the bilayer center in our simulations (Figure 12.3). As tryptophan fluorescence transition dipole is probably associated with the aromatic ring only, we now investigate the effect of having tryptophan instead of the wild type residues on the insertion depth measurements. To do so, we performed the M1W, F11W and F17W mutations in the Htt17_α_hrex wild type trajectories and optimized the side-chain orientation using SCWRL4 [252]. Of course, structural/topological alterations could be introduced by the mutations themselves, but simply taking into account the tryptophan side chain's geometry leads to 1.13 $\pm$ 0.09 nm (M1W), 0.93 $\pm$ 0.07 nm (F11W) and 1.32 $\pm$ 0.07 nm (F17W), which is in better agreement with the experimental measurements. Two supporting HREX simulations starting from a single α-helix at different insertion depths (2.0 and 2.5 nm from the bilayer's center, compared to 1.8 nm previously) converge to the same equilibrium values (Figure V.2).

The Htt17 membrane anchor has 6 residues that can form side chain hydrogen bonds as do-

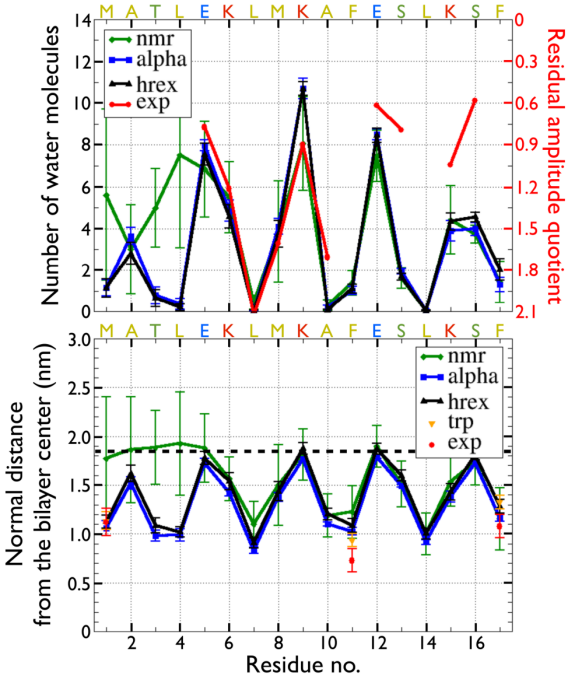


Figure 12.3 – Solvent accessibility and insertion depth of each residue of Htt17. Top panel : The solvent accessibility is quantified in terms of the number of water molecules within 0.35 nm from each residue (lhs axis). The rhs axis represents the residual amplitude quotient due to the addition of hydrophilic (Mn^{2+} ions) and hydrophobic (16-doxyl-stearic acid) paramagnetic relaxation reagents on Htt17 in the presence of DPC detergent micelles [318]. A larger residual amplitude quotient indicates a larger solvent accessibility of the side chain. Bottom panel : The insertion depth corresponds to the center-of-mass distance between each side chain and the bilayer along the direction perpendicular to the membrane surface (z-axis). The dotted line at 1.8 nm is the average position of the phosphorous atoms of the upper leaflet with respect to the center of the bilayer. The experimental values were measured using fluorescence quenching of Htt17 Trp mutants in the presence of POPC/POPS vesicles containing brominated POPC lipids [317]. The results for the MD simulations starting from the NMR model and a single α-helix are respectively shown by green diamonds and blue squares. The HREX simulation is shown by black triangles and the average position of the aromatic rings obtained from direct M1W, F11W and F17W mutations on the wild-type trajectory are shown by orange triangles. The value and the error bar for each residue are respectively the average and the standard deviation over all the averages on the 250–1000 ns time interval obtained from 11 independent MD simulations (Table 12.I). For HREX, the average and standard deviation are computed at the unscaled replica on the 50–250 ns time interval using 20-ns time windows. See Figure V.2 for the results on the two additional HREX simulations starting from the single α-helix at different insertion depths.

nors – Thr3, Lys6, Lys9, Ser13, Lys15 and Ser16 – with the oxygens of the phosphate group and the carboxyl groups of POPC. The average number and occurrence probability of H-bonds are shown in Table 12.III. Ordering these residues in terms of their propensity to form H-bonds yields Ser16 < Thr3 < Ser13 < Lys9 < Lys6 *lesssim* Lys15 in Htt17_nmr, Ser16 < Lys9 < Ser13

< Thr3 < Lys15 *lesssim* Lys6 in Htt17_α, and Ser16 < Lys9 < Thr3 < Ser13 < Lys15 < Lys6 in Htt17_α_hrex. For the two initial states, Ser16 shows the lowest H-bond occurrence probability with only 10%, while Lys6 and Lys15 form H-bonds more than 85% of the time. Two supporting HREX simulations starting from a single α-helix at different insertion depths (2.0 and 2.5 nm from the bilayer's center, compared to 1.8 nm previously) converge to a similar h-bond interaction pattern (Table V.II).

Table 12.III – Average number of hydrogen bonds and salt-bridges between Htt17 and the phospholipids of the membrane.

Interaction type	Residue	Htt17_nmr no. (%)	Htt17_α no. (%)	Htt17_hrex no. (%)
Hydrogen bonds	Thr–3	0.9 ± 0.3 (14 ± 19)	1.0 ± 0.0 (61 ± 30)	1.0 ± 0.0 (51 ± 30)
	Lys–6	1.8 ± 0.3 (85 ± 17)	1.9 ± 0.2 (92 ± 4)	2.2 ± 0.2 (95 ± 5)
	Lys–9	1.4 ± 0.2 (56 ± 19)	1.4 ± 0.2 (38 ± 11)	1.1 ± 0.1 (32 ± 9)
	Ser–13	1.0 ± 0.0 (45 ± 22)	1.0 ± 0.0 (54 ± 8)	1.0 ± 0.1 (54 ± 19)
	Lys–15	1.8 ± 0.3 (86 ± 8)	1.7 ± 0.1 (90 ± 4)	1.7 ± 0.2 (90 ± 3)
	Ser–16	1.0 ± 0.0 (9 ± 5)	1.0 ± 0.0 (6 ± 2)	1.0 ± 0.1 (5 ± 4)
Salt-bridges	Glu–5	0.9 ± 0.3 (1 ± 2)	1.0 ± 0.0 (1 ± 1)	1.0 ± 0.1 (1 ± 1)
	Glu–12	0.9 ± 0.3 (1 ± 1)	1.0 ± 0.0 (1 ± 0)	1.0 ± 0.0 (1 ± 1)
	Lys–6	1.1 ± 0.1 (49 ± 11)	1.2 ± 0.1 (60 ± 8)	1.3 ± 0.2 (65 ± 20)
	Lys–9	1.2 ± 0.1 (40 ± 13)	1.1 ± 0.0 (28 ± 5)	1.1 ± 0.1 (28 ± 7)
	Lys–15	1.1 ± 0.1 (51 ± 8)	1.1 ± 0.1 (48 ± 9)	1.0 ± 0.1 (38 ± 10)

For a given residue, statistics are performed only over the frames in which the residue forms at least one H-bond or salt-bridge. The numbers in parenthesis represent the probability of occurrence in percentage. The value and the error bar for each residue are respectively the average and the standard deviation over all the averages on the 250–1000 ns time interval obtained from 11 independent MD simulations (Tables V.II and V.III). For the HREX simulation, the average and standard deviation are computed at the unscaled replica on the 50–250 ns time interval using 20-ns time windows. See Tables V.II and V.III for the results on the two additional HREX simulations starting from the single α-helix at different insertion depths.

The positively charged lysines of Htt17 also form salt-bridges with the negatively charged phosphate group of the phospholipids. With occurrence probabilities ranging from 28 to 60%, we see that such salt-bridges are a key component of the interactions between Htt17 and the bilayer (Table 12.III). Moreover, these residues can participate in up to three salt-bridges with different phospholipids. In order of occurrence probability, we find Lys9 < Lys6 *lesssim* Lys15 in Htt17_nmr, while Lys9 < Lys15 < Lys6 for the single α-helix in both Htt17_α and Htt17_α_hrex. While Lys9 has a lower probability to form salt-bridges with the phospholipids, it often forms intramolecular salt-bridges with either Glu5 and Glu12 contrary to the other lysines (Table V.IV). In principle, the negatively charged glutamic acids of Htt17 could form salt-

bridges with the positively charged nitrogen of the choline group. The occurrence probability for such salt-bridges is however very low (1–2%) as the three methyl groups surrounding the nitrogen cause steric hinderance to the salt-bridge formation (Table 12.III). Two supporting HREX simulations starting from a single α-helix at different insertion depths (2.0 and 2.5 nm from the bilayer's center, compared to 1.8 nm previously) converge to a similar salt-bridge interaction pattern (Table V.III).

12.5 Discussion

The first 17 amino acids at the amino-terminus (Htt17) of huntingtin, a large protein related to Huntington's disease and with multiple biological functions [26, 168, 358, 535, 539], modulates its localization, function, aggregation and degradation in the cell [4, 14, 16, 175, 302, 398, 443, 459, 463, 536]. For instance, the amphipathic character of the Htt17 segment is responsible for huntingtin's membrane anchoring properties and regulates its interaction with specific membrane-containing organelles of the cell [16]. Recently, solution NMR has been used to resolve the structure of Htt17 in the presence of DPC detergent micelles [319]. Solid-state NMR further unveiled its orientation on a POPC bilayer using ^{15}N chemical shift and ^{2}H quadrupolar splitting measurements [318].

The molecular dynamics (MD) and Hamiltonian replica exchange (HREX) simulations described in the present paper are combined with these NMR measurements to yield the first high-resolution model for the membrane anchoring properties of the Htt17 segment of huntingtin. Our results refine the atomistic picture of Htt17 on a phospholipid bilayer in terms of its structure, orientation, key membrane interactions (hydrogen bonds, salt-bridges, insertion depth and solvent accessibility) and membrane perturbations (order parameter, area per lipid and thickness), as well as the influence of the phospholipid type on those observations. We now discuss each finding and their implications in details.

12.5.1 The Htt17 monomer forms a stable α-helix on a POPC bilayer

In aqueous solution, the Htt17 monomer is mostly disordered as shown by solution NMR [459], while CD spectroscopy suggests that it can form transient helical structures [16, 319, 459, 504]. Transition to a stable α-helix occurs upon self-association [209] as well as upon addition of phospholipid vesicles, detergent micelles or apolar compounds to an aqueous buffer [16, 319,

459]. Recently, the structure of Htt17 in the presence of DPC micelles has been resolved using solution NMR spectroscopy : it is disordered from residues 1 to 5 and it is an α-helix for the rest of the sequence [319]. Despite indirect indications from ssNMR [318], it had not be verified, yet, whether Htt17 adopts exactly the same structure on a phospholipid bilayer.

Our simulations extend previous observations by showing, with many independent MD and HREX simulations, that the first residues of Htt17 also form a stable α-helix on a POPC bilayer (Figure 12.1). Moreover, the ^{15}N chemical shifts of Leu7, Phe11 and Phe17 sampled during those simulations agree with solid-state NMR measurements of Htt17 on a POPC bilayer [318] (Table 12.II). These values are consistent with an in-plane orientation of the amphipathic Htt17 with respect to the membrane surface as depicted in Figure 12.2. Our observations are further supported by very similar results obtained in terms of structure (Figure V.4) and chemical shifts (Table V.I) using a different forcefield (AMBER99sb*-ILDN/SLIPIDS, Table 12.I).

The increased stability of the Htt17 monomer as an α-helix due to the presence of a phospholipid bilayer could have a strong impact on its oligomerization. In aqueous solution, Htt17 is mostly disordered at the monomeric level as shown by solution NMR experiment [459], but it forms stable tetrameric bundle of four α-helical Htt17 [209]. When linked to the polyglutamine segment (Q_N) that follows Htt17 in huntingtin's sequence, the aggregation pathway of Q_N is significantly modified [209, 210]. More precisely, Htt17Q_N has two main aggregation pathways in direct kinetic competition : (i) one is initiated by the formation of α-helical tetrameric bundles of Htt17 that combine to form larger oligomeric assemblies favoring Q_N fibrillation due to an increase in its local concentration, and (ii) the other is independent of structure formation in Htt17 resulting in an aggregation pathway that is very similar to Q_N alone [210]. The first pathway yields faster fibrillation kinetics [275, 459].

Here, we observe that the Htt17 monomer forms a stable α-helix on the bilayer (Figure 12.1) that is significantly different from its mostly disordered structure in aqueous solution as observed from other simulations using the same forcefield [111]. The presence of the membrane favors the formation of α-helical structures in Htt17 that could then ease the nucleation of the α-helical tetrameric bundle of Htt17. Consequently, the oligomerization and fibrillation kinetics of Htt17Q_N could be enhanced by the presence of a phospholipid bilayer [319] as for other amyloidogenic proteins [263]. On the membrane, as observed from MD simulations, the Q_N region of Htt17Q_N and Htt17$Q_N P_{11}$ is indeed easily available for aggregation as it lies on the surface of

the phospholipid heads [111, 335] and weakly interact with Htt17 [111]. The converse argument is also possible : the formation of oligomeric structures containing an α-helical Htt17 prior to its binding on the phospholipid bilayer could favor its insertion. Indeed, previous simulations suggest that the complete partitioning of Htt17 non-polar residues is eased by the presence of α-helical structures prior to binding, while the reordering of disordered structures on the bilayer are impeded by strong electrostatic interactions between Htt17 and the phospholipids [111].

12.5.2 The key residues for Htt17–membrane interaction are crucial to huntingtin's function

The Htt17 segment of huntingtin is involved in several posttranslational modifications such as SUMOylation [443] and phosphorylation [4, 14, 175, 463] that are crucial to its physiological and pathological functions. More precisely, SUMOylation implicating Lys6 and Lys9 correlates with increased neurodegeneration in Drosophila possibly through the modulation of huntingtin localization and aggregation [443]. Phosphorylation of Ser13 and Ser16 increases huntingtin clearance, reduces the toxicity in a mouse model, regulates other posttranslational modifications of Htt17, and increases nuclear localization of huntingtin [14, 175, 463]. Phosphorylation of Thr3 increases the formation of insoluble aggregates and a phosphomimetic mutation of the threonine to an aspartic acid reduces the neurodegeneration in Drosophila [4]. Mutations or truncation of the non-polar residues implicated in the nuclear export signal (NES) of Htt17 – L4, L7, F11 and L14 – leads to a significant increase of huntingtin accumulation in the nucleus [16, 104, 536]. Neutralization of the lysines and glutamic acids by substitution to alanines modulates the membrane composition targeted by Htt17 [16].

In our simulations, we observe that many of these crucial residues are involved in specific interactions with the phospholipids. For instance, salt-bridges between Lys6 and Lys15 and the phospholipids occur often, while Lys9 is mainly involved in intramolecular salt-bridges with either of the two glutamic acids in Htt17 (Tables 12.III and V.IV). These charged residues as well as Ser13 also regularly form hydrogen bonds with the phospholipids (Table 12.III). Moreover, non-polar sequestration of L7, F11 and L14 inside the hydrophobic core of the membrane is crucial to the anchoring of Htt17 (Figure 12.3). Our observations are reinforced by similar results in terms of solvent accessibility (Figure V.5), insertion depth (Figure V.6), hydrogen bonds (Table V.II) and salt-bridges (Table V.III) using a different forcefield (AMBER99sb*-

ILDN/SLIPIDS, Table 12.I). Our simulations therefore suggest that perturbation of these key membrane interactions (Table 12.III and Figure 12.3) will likely impact Htt17's ability for membrane binding in various ways consequently affecting huntingtin localization in the cell and potentially its function.

12.5.3 The network of electrostatic interactions depends on the phospholipid type, but not the configuration of Htt17

The structure and orientation of Htt17 remain almost unchanged when inserted in membranes of different composition as shown from solution NMR, solid-state NMR and CD spectroscopy [319]. In contrast, the ability of this membrane anchor to target phospholipid vesicles highly depends in the membrane composition as shown from the binding affinity extrapolated from CD measurements and tryptophan-emission fluorescence [317, 319]. More precisely, Htt17 binds more favorably to membranes containing anionic phospholipids such as PS and PG than to the zwitterionic PC, while the presence of cholesterol has the opposite effect. A similar behavior has also been observed for the exon 1 of huntingtin [230].

We can compare our results on a POPC bilayer to a previous study in which we investigated Htt17 on a POPE bilayer using MD simulations [111]. As these simulations were done using the AMBER99sb*-ILDN/SLIPIDS forcefield, we compare them to the results on the POPC bilayer obtained using the same forcefield (Table 12.I). In each case, we have two simulations of 500 ns each for POPE and of 1000 ns each for POPC starting from different velocity distributions. Htt17 is initially a single α-helix inserted in the bilayer below the phosphate group of the phospholipids.

We observe that Htt17's structure (Figure V.7) and orientation (Table V.V) are quantitively similar in the zwitterionic POPC and POPE bilayers. In contrast, the insertion depth of Htt17 (Figure V.8) as well as its network of salt-bridges and hydrogen bonds with the phospholipids (Table V.VI) are different. More precisely, the main difference is that the glutamic acids (Glu5 and Glu12) form salt-bridges and hydrogen bonds with the amine group of PE, while such interactions are not present for the choline group of PC. This shifts Htt17 nearer from the phospholipid heads without changing significantly its solvent accessibility. In the previous simulations on POPE, we also observed that electrostatic interactions initialize the binding of Htt17 on the phospholipid bilayer [111].

Taken together, experiments and simulations suggest that the main interaction responsible for the structure and orientation of Htt17 is the sequestration of its non-polar residues in the hydrophobic core of the membrane because this is independent of the phospholipid species. In contrast, the binding affinity of Htt17 on phospholipid bilayers appears to depend on specific electrostatic interactions that are modulated by the membrane composition. The main role of the charged residues could be to target selectively membrane with specific phospholipid compositions. Indeed, experiment shows that mutating the charged residues to alanines can change the localization of huntingtin in the cell [16].

Beside the phospholipid type, the presence of cholesterol also has an impact : increasing its concentration significantly impedes Htt17's ability to bind and permeabilize phospholipid membranes [317, 319]. As discussed in the next section, our simulations show that Htt17 induces local perturbations in the physical properties of the bilayer such as membrane thinning, reduced phospholipid area and decreased phospholipid order parameters (Figure 12.4). These perturbations are due to the neighboring phospholipids of Htt17 extending their acyl chains to cover its non-polar surface (Figure 12.4). Experiment show that the effect of cholesterol is the opposite : it increases the order of the phospholipid acyl chains and the membrane thickness. Doing so, the cholesterol might protect the membrane from accommodating the binding state of the Htt17 monomer thereby reducing its binding affinity and further oligomerization in the membrane.

12.5.4 The Htt17 monomer locally perturbs the physical properties of the bilayer

Experiments show that the presence of 2.5 mole % of Htt17 on POPC, POPE/POPG and POPC/POPS bilayers significantly reduces the order parameters of the phospholipid palmitic chain [319]. Few μM of Htt17 are sufficient to induce leakage of large POPC and POPC/POPS unilamellar vesicles [319]. Globular aggregates of the exon 1 of huntingtin increase, similarly to other amyloid protein [73], the roughness and change the mechanical properties of total brain lipid extract (TBLE) bilayers as shown by atomic force microscopy (AFM) [71].

At the monomer level, we observe that the presence of Htt17 on a POPC membrane already leads to local perturbations of the bilayer physical properties as summarized in Figure 12.4. For instance, the area per lipid and membrane thickness are decreased within a distance of about 1 nm from Htt17 with respect to the bulk phospholipids. This reordering of the bilayer accommodates Htt17 by allowing its neighboring phospholipids to extend their acyl chain to

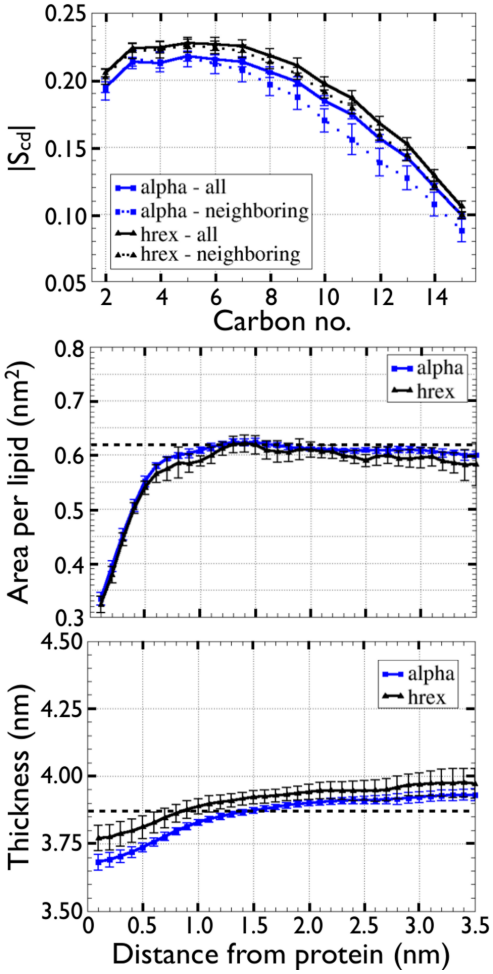


Figure 12.4 – Perturbations of the bilayer properties by Htt17. Results for Htt17_α and Htt17_α_hrex are respectively shown in blue squares and black triangles. (A) Average order parameters of the palmitic chain carbons of the neighboring (dotted) and all (line) phospholipids. The neighbouring phospholipids are defined as those within 1.0 nm of Htt17 in the upper leaflet. (B) Average area per phospholipid computed using VTMC [329] as a function a their distance from Htt17 and (C) average membrane thickness computed using GridMAT [7]. For (B) and (C), the black dotted line represents the averaged value for a simulated POPC bilayer without Htt17. The value and the error bar for each residue are respectively the average and the standard deviation over all the averages on the 250–1000 ns time interval obtained from 11 independent MD simulations (Table 12.I). For HREX, the average and standard deviation are computed at the unscaled replica on the 50–250 ns time interval using 20-ns time windows. See Figure V.2 for the results on the two additional HREX simulations starting from the single α-helix at different insertion depths.

cover its non-polar surface as depicted by Figure 12.2. The resulting order parameters of these acyl chains is decreased with respect to the membrane bulk (Figure 12.4) in agreement with other MD simulation of a disordered KKHtt17Q_{35}KK on a DPPC bilayer [335] and with experimental

measurements on a POPC bilayer with 2.5% mole of Htt17 [319].

Theoretical studies show that such local membrane deformations around a peptide promote dimerization of inserted amphipathic peptides on phospholipid membranes [193, 521, 522]. This could favor the dimerization and further oligomerization of Htt17. Dimerization could be initiated and stabilized by the formation of electrostatic interactions as proposed from ssNMR orientation measurements on Htt17 [319]. Our simulations indeed show that most of the charged residues stay accessible to the solvent (Figure 12.3). We expect the rate limiting step of such process to be the need to break the strong peptide–phospholipids interactions (Table 12.III) as some of the neighboring phospholipids of Htt17 exchange very slowly with the membrane bulk : about 10 phospholipids stay in contact with the peptide for more than 400 ns and 20 for more than 150 ns (Figure V.9).

Following dimerization, Htt17 could then form, as observed experimentally, larger globular aggregates that significantly change the membrane properties [71] and cause vesicle permeabilization [318, 319]. The preferred mechanism of permeabilization by Htt17 – pore formation, carpet model or detergent model [505] – remains however unknown. While pore formation of Htt17 alone is probably not relevant for the biological activities as it is weak in membranes with cholesterol or absent in POPE/POPG membrane [319], it could be different in the presence of the amyloidogenic Q_N region (Htt17Q_N). Further integrated experimental and computational studies on this front will be essential to compare huntingtin with the other amyloid proteins that are known to perturb phospholipid membranes [264, 525]. While the Q_N region can perturb on its own the membrane integrity [186, 219, 327], we suggest from our results on the Htt17 and Htt17Q_N monomers [111] and previous experimental observations a new paradigm : the role of the amyloidogenic Q_N region could be to stabilize, in a length-dependent manner, the oligomeric assemblies of huntingtin exon 1 as it stays on the surface and is easily accessible for oligomerization. The amphipathic Htt17, for its part, could play a fundamental role by initiating the binding and insertion of the monomer in the membrane as well as by perturbing the physical properties of the bilayer in a manner to favor its oligomerization, similarly to other α-helical amphipathic peptides [31, 33, 505].

12.6 Conclusions

The function and localization of huntingtin are intrinsically dependent on the first 17 amino acids at its amino-terminus (Htt17). This amphipathic segment serves as a membrane anchor and is situated just before the amyloidogenic polyglutamine segment of huntingtin. In this study, we present a high-resolution atomistic model of the Htt17 monomer on a phospholipid bilayer by combining atomistic molecular dynamics (MD) and Hamiltonian replica exchange (HREX) simulations together with experimental results from solution and solid-state nuclear magnetic resonance, as well as fluorescence spectroscopy. Our model quantifies the key interactions between Htt17 and the phospholipids of the bilayer responsible for its structure, orientation and insertion depth, as well as the resulting physical perturbation of the bilayer structure. It rationalizes, at the atomic level, the potential effect of different membrane compositions and posttranslational modifications of Htt17 on its ability to target phospholipid bilayers. It also describes the impact of the monomer configuration and membrane perturbations on Htt17 and Htt17Q_N self-association into oligomeric complexes of potential relation to the pathogenesis of huntingtin. Finally, we observe that the sampling enhancing method HREX, which enables the system to easily move out of local minima that could bias the sampling, is ideal for peptide–membrane system as it is performed at room temperature, as it does not perturb the membrane physical properties and as it speeds up the convergence of the observables. This relatively inexpensive computational method could be readily applicable to investigate a wider variety of membrane compositions to precisely unveil, at the atomic level, the concentration dependence of cholesterol and charged phospholipids on Htt17, Htt17Q_N and Htt17$Q_N P_{11}$ structure, orientation, key interactions, binding affinity and oligomerization on the surface of phospholipid membranes.

12.7 Supporting Material

A supplement to this article can be found in Appendix V.

12.8 Author contributions

S.C., B.B., and N.M. designed the research ; S.C. and V.B. performed the research ; S.C., V.B., and E.S.S. contributed analytic tools ; S.C., V.B., E.S.S., B.B., and N.M. analyzed the data ; and S.C., B.B., and N.M. wrote the manuscript.

12.9 Acknowledgments

The authors thank Giovanni Bussi for his help in setting up the Hamiltonian replica exchange (HREX) simulation.

Computations were made on the computing facilities of Calcul Québec/Compute Canada. The operation of these supercomputers is funded by the Canada Foundation for Innovation, Nano Québec, RMGA, and the Fonds de Recherche Québécois sur la Nature et les Technologies.

This work was funded in part by the Canada Research Chairs program, the Fonds de Recherche Québécois sur la Nature et les Technologies, the Natural Sciences and Engineering Research Council of Canada, and the Fonds de Recherche en Santé du Québec, the Agence Nationale de la Recherche (projects TRANSPEP, ProLinIn, membraneDNP, and the LabEx Chemistry of Complex System), the University of Strasbourg, the Centre National de la Recherche Scientifique, the American Foundation for Research on Huntington's Disease (Cure Huntington's Disease Initiative), and Les Réseaux Thématiques de Recherche Avancée International Center of Frontier Research in Chemistry.

CHAPITRE 13

OPTIMISATION DES CHAMPS DE FORCE GROS-GRAINS AAOPEP ET SOPEP

Refining PEP-FOLD for *de novo* peptide structure prediction with an all-atom extension of the coarse-grained potential OPEP

Sébastien Côté[†], Vincent Binette[†], Pierre Tufféry[§], and Normand Mousseau[†]

† Département de Physique and Groupe d'étude des protéines membranaires (GÉPROM), Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec), Canada

§ Université Paris Diderot, Sorbonne Paris Cité, INSERM U973, MTi, 75205 Paris, France

Manuscrit en préparation. La validation de la paramétrisation de sOPEP et aaOPEP se poursuivra dans les mois à venir par d'autres étudiants des groupes des professeurs Normand Mousseau et Pierre Tufféry. Cette section a été écrite dans l'optique d'une publication future.

13.1 Abstract

PEP-FOLD is a web server that predicts the three dimensional structure of peptides of less than 50 amino acids given their sequence. The predicted three dimensional structures are ranked according to an energy scoring function (sOPEP) derived from the coarse-grained potential OPEP that has been used to investigate numerous problematics of interest ranging from protein folding to amyloid protein aggregation. Here, to refine PEP-FOLD structure predictions, we report on the re-optimization of the sOPEP scoring function as well as the development and optimization of an extension of OPEP core philosophy to the all-atom regime. First, the parameter optimization of sOPEP is done on a training set consisting of structural decoys generated for peptides with various α, β and α/β supersecondary structure motifs. The optimization is performed by a genetic algorithm that improves the parameters of the scoring function so that it can discriminate between native-like, near-native and non-native folds. Second, we evaluate the discrimination sensitivity of this re-optimized version on a totally separate validation set. All in all, our training and validation sets count many thousands of structural decoys and they respectively

consist of 50 and 105 peptides having less than 70 amino acids. Third, building on the success of OPEP, we develop an all-atom extension (aaOPEP) of its core philosophy : (i) solvent- and charge-free, (ii) specific non-bonded parameters for each side-chain/side-chain pair to take into account their hydrophobic interactions, as well as (iii) explicit and cooperative H-bonds. We further add explicit desolvation barriers as well as explicit side-chain/side-chain salt-bridges and H-bonds. The parameter optimization of aaOPEP follows the same protocol as for sOPEP. Finally, we evaluate the ability of aaOPEP to refine the best structure predictions of PEP-FOLD with the use of restrained molecular dynamics simulations and energy minimizations.

13.2 Introduction

The structure prediction of a protein directly from its sequence of amino acids is a long standing goal in structural biology [17, 136, 238, 267, 330, 331]. A wealth of computational methods ranging from brute force physics-based to statistical bioinformatic-based approaches were developed over the years [100]. These latter are particularly attracting given their prediction speed and considering the enormous number of protein-coding sequences that are unveiled by the genome projects. These methods are either de novo or template-based [158, 530]. While template-based methods use sequence homology between the target sequence having an unknown fold to the sequence of proteins having a known fold, de novo methods work even if the target sequence has no near homologs with a known fold. The flexibility of de novo methods allows the design of completely new folds [247, 253] that could be used in biotechnological applications [237].

Nowadays, the combination of de novo and physics-based approaches into hybrid methods is deemed more and more essential [236, 330]. For instance, the PEP-FOLD method for de novo peptide and miniprotein structure prediction combines a fragment assembly method [309, 310, 430, 462] and a scoring function (sOPEP) based upon a coarse-grained potential called OPEP [310, 311].

The fragment assembly method of PEP-FOLD uses a structural alphabet (SA) of 27 fragment types that are 4-residue long [309, 310, 430, 462]. The SA letters represent the local conformations of 4-residue long fragments that optimally encode the structural diversity observed in the Protein Data Bank. A support vector machine (SVM) has been trained to predict the probability of each SA letter at all positions in the sequence. A greedy superposition of the probabilistically chosen SA letters is then performed. During the reconstruction, the sOPEP scoring function is

used to evaluate the best partial folds. Incorrect folds are discarded in order to explore more efficiently the conformational space accessible to the sequence of amino acids. Once the reconstruction is completed, the best folds are determined using clustering and sOPEP.

The sOPEP scoring function is directly based upon the OPEP potential that has been applied on numerous problematics of interest ranging from protein folding to amyloid protein aggregation using a variety of computational methods such as Monte Carlo (MC), the activation-relaxation technique (ART), molecular dynamics (MD), replica-exchange molecular dynamics (REMD), simulated tempering (ST) and metadynamics (MetaD) [444]. This coarse-grained model represents each amino acid with all their backbone atoms and only one bead for their side-chain atoms. The core philosophy of OPEP is based upon the following characteristics : it is solvent-free and charge-free, it has specific side-chain/side-chain interaction types that are finely tuned against protein structure and thermodynamics, and it has explicit, directional and cooperative hydrogen bonds [311].

Recently, the performance of the structure prediction of PEP-FOLD has been improved by exploring more efficiently the SA letter sequence space accessible to the protein [430]. It now compares well with other programs such as the state-of-the-art ROSETTA [118] for peptides and miniproteins (< 52 amino acids). It is able to generate near-native or native models for 95% of 56 structurally diverse peptides with 25 to 52 amino acids without any coarse-grained and all-atom refinement steps. Moreover, near-native or native models are among the five best sOPEP score for 80% of the peptides. In light of these results, there are two other areas for which the efficiency of PEP-FOLD could be further increased : (i) the sOPEP energy function needs to be more discriminative and (ii) the presence of a final all-atom refinement step to improve the quality of the structural predictions is needed.

In this article, we re-optimize the sOPEP scoring function a genetic algorithm and following the same protocol as before [310, 311] with a more extensive training set. The new set consists of 50 proteins having between 51 and 70 amino acids that have more complex folds that better cover the whole spectrum of native side-chain/side-chain interaction types compared to the previous optimization. The decoys generated for each protein also better cover the spectrum of structural diversity with respect to the native state. Moreover, we put more emphasis on the discrimination of the native-like and near-native states.

We also develop and optimize an all-atom extension of the core philosophy of OPEP. This

new physics-based forcefield, aaOPEP, is used at the end of the PEP-FOLD method to refine the best fold candidates in a representation that is not too far from OPEP – solvent-free and charge-free, specific side-chain/side-chain interactions as well as explicit, directional and cooperative hydrogen bonds – at the same time as having the atomic representation that is necessary for a better packing of the side-chains. Indeed, it was noted previously that the core packing of OPEP is not ideal, which is normal given that it is a coarse-grained model, in some globular proteins [142] and loops [441]. This higher resolution model also allows main-chain/side-chain H-bonds that are often seen in peptides and miniproteins. The differentiability of the aaOPEP forcefield allows the use of restrained molecular dynamics simulations and energy minimizations to refine the best structure predictions of PEP-FOLD.

13.3 Material and Methods

In this study, we re-optimize and validate the sOPEP scoring function using the same approach as before, but with a more extensive training and validation sets than previously used [310, 311]. We also develop and optimize the aaOPEP potential to refine PEP-FOLD structure predictions [309, 310, 430, 462] by means of restrained molecular dynamics simulations and energy minimization. Below, we present the sOPEP and aaOPEP energy functions and the parameters that are optimized. Then, we describe the choice of peptides for the training and validation sets as well as the protocols used to generate the structural decoys for all chosen peptides. Finally, we present the optimization and validation procedures as well as the metric used to quantify the structural similarity between different folds of the same peptide.

The sOPEP scoring function. The sOPEP scoring function is directly derived from the OPEP coarse-grained potential that is a six-bead model with only one bead for the side-chain (N, H_N, C_α, C, O and SC) [310, 311]. Basically, sOPEP uses a different function for the side-chain/side-chain interactions to avoid non-natural clashes by controlling the distance at which the energy is zero. This distance is unique for each pair type and it is set to the 0.1 quantile of the distance distribution of 2 248 structures from the Protein Data Bank with less than 30% sequence identity. Moreover, sOPEP in PEP-FOLD does not have the bonded terms of OPEP. These latter terms are included back for performing the energy minimizations and molecular dynamics simulations in our study. All other energy terms are as in OPEP. The number of weights of sOPEP is 217 : 1 weight for the 1–4 Lennard-Jones interactions and 1 for all others, 1 weight for all

C_α–C_α interactions, 210 weights for all possible SC–SC pair types (20 types of natural amino acid), 1 weight for i/i+4 H-bond and 1 weight for all other H-bond types, as well as 1 weight for α-helix cooperativity and 1 weight for β-sheet cooperativity. We refer to our previous works for more detail [310, 311].

The aaOPEP potential. The functional form of the all-atom OPEP (aaOPEP) potential is differentiable to allow the fast minimization of the *de novo* structure prediction of PEP-FOLD and their refinement using restrained molecular dynamics. The aaOPEP potential has the following form :

$$\mathrm{E} = \mathrm{E_{bond}} + \mathrm{E_{angle}} + \mathrm{E_{tors}} + \mathrm{E_{itors}} + \mathrm{E_{LJ}} + \mathrm{E_{SB}} + \mathrm{E_{SC\text{–}SC}} + \mathrm{E_{C\alpha\text{–}C\alpha}} + \mathrm{E_{HB}} + \mathrm{E_{HB2}} \tag{13.1}$$

where the total energy is grouped into two main terms. The bonded interactions including bond stretching ($\mathrm{E_{bond}}$), bond angle bending ($\mathrm{E_{angle}}$), torsion angle rotation ($\mathrm{E_{tors}}$) and improper torsion angles ($\mathrm{E_{itors}}$). The non-bonded interactions include Lennard-Jones interactions ($\mathrm{E_{LJ}}$), explicit salt-bridges ($\mathrm{E_{SB}}$), specific Lennard-Jones interactions between the side-chains ($\mathrm{E_{SC\text{–}SC}}$), special interactions between the carbon-alpha ($\mathrm{E_{C\alpha\text{–}C\alpha}}$), explicit hydrogen bonds ($\mathrm{E_{HB}}$) and cooperativity between the main-chain hydrogen bonds ($\mathrm{E_{HB2}}$). We now present the functional form of each term.

The bonded interactions are expressed by

$$\mathrm{E_{bond}} = \sum_{\{i,j\} \in \text{ bonds}} \frac{1}{2} k^r_{ij} \left(r_{ij} - r^{\mathrm{eq}}_{ij} \right)^2 \tag{13.2}$$

$$\mathrm{E_{angle}} = \sum_{\{i,j,k\} \in \text{ angles}} \frac{1}{2} k^\theta_{ijk} \left(\theta_{ijk} - \theta^{\mathrm{eq}}_{ijk} \right)^2 \tag{13.3}$$

$$\mathrm{E_{tors}} = \sum_{\{i,j,k,l\} \in \text{ torsions}} \left[\sum_{n_{ijkl}} k^\phi_{ijkl,n} \left(1 - \cos\left(n\phi_{ijkl} - \delta_{ijkl,n} \right) \right) \right] \tag{13.4}$$

where $\mathrm{E_{bond}}$ is a sum over all pairs of atoms (ij) linked by a covalent bond with each term depending on three parameters that are the recoil constant (k^r_{ij}), the equilibrium length (r^{eq}_{ij}) and the actual length (r_{ij}) of the current covalent bond, $\mathrm{E_{angle}}$ is a sum over all atom triplets (ijk) linked by two consecutive covalent bonds with each term depending on three parameters that are

the recoil constant (k^{θ}_{ijk}), the equilibrium angle ($\theta^{\mathrm{eq}}_{ijk}$) and the actual angle ($\theta_{ijk}$) of the current bond angle, and E_{tors} is a sum over all atom quartets (ijkl) linked by three consecutive covalent bonds with each term being a sum over a few Fourier terms depending on four parameters that are the energy amplitude (k^{ϕ}_{ijkl}), the angular frequency (n), the phase shift ($\delta_{n,ijkl}$) and the actual torsion angle (ϕ_{ijkl}). E_{itors} has the same functional form as E_{tors}, but it is instead used to enforce geometrical features between specific atom quartets such as planar and tetrahedral geometries as well as the chirality of the peptide plane.

The parameters for the bonded terms are taken from the AMBER99sb*-ILDN forcefield [382] that is recognized as one of the best forcefield for studying protein folding and stability in explicit solvent [28, 98, 279, 380, 383]. Bonds and bond angles are at the heart of the polypeptide chain geometry and their parameters were derived from X-ray crystal structures and normal mode analysis [103]. Torsion angles originate from quantum mechanical interactions between the quartet of atoms linked by consecutive covalent bonds and their parameters were determined using *ab initio* computations [490]. Compared to AMBER99sb [190], the parameters of some torsion angles were recently changed to improve the relative stability between helix and coil structures through the re-parametrization of the ψ torsion angle (AMBER99sb*) [46], and to improve the energy landscape of the χ_1 and χ_2 torsion angles for the isoleucine, leucine, aspartate and asparagine (AMBER99sb-ILDN) [281]. In summary, we keep the well-parametrized skeleton of the polypeptide chain and we focus on the parametrization of the non-bonded interactions specific to aaOPEP. This is sufficient for the moment as aaOPEP is used to refine PEP-FOLD structure predictions using short molecular dynamics simulations with restraints on the ϕ and ψ dihedral angles. Adjustments to the bonded parameters could eventually be easily added, if necessary, to our potential without changing the non-bonded parameters as done a number of times for the AMBER forcefield. We will ensure the correctness aaOPEP to model the dynamics of protein folding and aggregation in future works.

The non-bonded interactions are expressed in part by

$$\mathrm{E_{LJ}} = \sum_{\substack{\forall\{i,j\} \notin \\ \mathrm{SC\text{–}SC} \cup \mathrm{C\alpha\text{–}C\alpha}}} \varepsilon_{ij} \left[\left(\frac{r^0_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r^0_{ij}}{r_{ij}} \right)^{6} \right] \tag{13.5}$$

$$\mathrm{E_{SC\text{–}SC}} = \sum_{\{i,j\} \in \mathrm{SC\text{–}SC}} \mathrm{w}^{\mathrm{SC\text{–}SC}}_{ij} \varepsilon_{ij} \left[\left(\frac{r^0_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r^0_{ij}}{r_{ij}} \right)^{6} \right] \tag{13.6}$$

$$\mathrm{E_{C\alpha\text{–}C\alpha}} = \sum_{\{i,j\} \in \mathrm{C\alpha\text{–}C\alpha}} \mathrm{w}^{\mathrm{C\alpha\text{–}C\alpha}} \mathrm{U_{DB}} \left(r_{ij} \right) \tag{13.7}$$

where $\mathrm{E_{LJ}}$ is the standard Lennard-Jones interaction described by two parameters, which depend on the atom type of each atom, that are the minimum energy (ε_{ij}) and the distance (r^0_{ij}) at which it occurs. These parameters are taken from AMBER99sb*-ILDN and were derived from density and enthalpy of vaporization of various organic liquids [103]. For $\mathrm{E_{LJ}}$, the sum is over all pairs of atoms (ij) ; except between two carbon alpha and two aliphatic or aromatic carbon atoms in the side chain as these are respectively taken into account by $\mathrm{E_{C\alpha\text{–}C\alpha}}$ and $\mathrm{E_{SC\text{–}SC}}$. For $\mathrm{E_{SC\text{–}SC}}$, the standard Lennard-Jones interaction between aliphatic or aromatic carbon atoms is supplemented in aaOPEP by a weight depending on the pair of side chain type ($\mathrm{w}^{\mathrm{SC\text{–}SC}}_{ij}$). The role of this weight is to implicitly take into account the repulsion, dispersion, charged and solvation interactions between each side chain pair in a manner similar to the coarse-grained OPEP potential [311]. For neighboring side-chains (i/i+1), this weight is set to one. For $\mathrm{E_{C\alpha\text{–}C\alpha}}$, we model the interactions between carbon alpha by an attractive–repulsive potential with a desolvation barrier ($\mathrm{U_{DB}}$) following the previous works of Hue Sun Chan and co-workers on Gō models [95, 226, 286] :

$$\mathrm{U_{DB}}(r) = \begin{cases} \varepsilon Z(r) \left[Z(r) - 2 \right] & \text{for } r < r_{cm} \\ CY(r)^n \left[Y(r)^n/2 - \left(r_{db} - r^{2n}_{cm} \right) \right] /2n + \varepsilon_{db} & \text{for } r_{cm} \leq r < r_{db} \\ -B \left[Y(r) - h_1 \right] / \left[Y(r)^m + h_2 \right] & \text{for } r \geq r_{db} \end{cases} \tag{13.8}$$

with

$$
\begin{aligned}
Z(r) &= (r_{cm}/r)^{k} \\
Y(r) &= (r - r_{db})^{2} \\
C &= 4n\,(\varepsilon + \varepsilon_{db}) \,/\, (r_{db} - r_{cm})^{4n} \\
B &= m\varepsilon_{ssm}\,(r_{ssm} - r_{db})^{2(m-1)} \\
h_1 &= (1 - 1/m)\,(r_{ssm} - r_{db})^{2} \,/\, (\varepsilon_{ssm}/\varepsilon_{db} + 1) \\
h_2 &= (m-1)\,(r_{ssm} - r_{db})^{2m} \,/\, (1 + \varepsilon_{db}/\varepsilon_{ssm})
\end{aligned}
\tag{13.9}
$$

where r is the distance between the two atoms, r_{cm} is the position of the deepest energy minima ($-\varepsilon$), r_{db} is the position of the energy barrier (ε_{db}), r_{ssm} is the position of the second energy minima (-ε_{ssm}), $n \in \mathbb{N}$ and $m \in \mathbb{N}$. Following the seminal works of Hue Sun Chan and collaborators [84, 226], we let $r_{ssm} = r_{cm} + 0.3$ nm, $r_{db} = (r_{ssm} + r_{cm})/2$, $\varepsilon_{db} = 0.1\varepsilon$, $\varepsilon_{ssm} = 0.2\varepsilon$ $k = 6$, $m = 3$ and $n = 2$. The desolvation barrier is known to be crucial for small peptides with two-state-like kinetics and thermodynamics as the experimentally observed enthalpic barriers to folding [227, 285, 286], folding cooperativity [227, 228, 286] and diversity of folding rates [154, 228] originate in part from this. Moreover, it allows the native state to be structurally more specific, which will increase the sensitivity of the structure prediction of PEP-FOLD. In our optimization protocol, we calibrate the pre-factor $\mathrm{w}^{\mathrm{C}\alpha\text{–}\mathrm{C}\alpha}$ of $\mathrm{E}_{\mathrm{C}\alpha\text{–}\mathrm{C}\alpha}$. These are only applied for C_α separated by more than 3 residues (j > i+3).

The non-bonded interactions also contain

$$
\mathrm{E}_{\mathrm{SB}} = \sum_{\{i,j\} \in \mathrm{SB}} \mathrm{w}^{\mathrm{SB}} \mathrm{U}_{\mathrm{DB}}\left(r_{ij}\right) \tag{13.10}
$$

$$
\mathrm{E}_{\mathrm{HB}} = \sum_{\substack{\forall\{i,j\} \in \\ \mathrm{ACC–H}}} w_{ij}^{\mathrm{HB}}\, \varepsilon_{ij}^{\mathrm{HB}} \left[5 \left(\frac{r_{\mathrm{HB}}^{0}}{r_{ij}} \right)^{12} - 6 \left(\frac{r_{\mathrm{HB}}^{0}}{r_{ij}} \right)^{10} \right] \cos^{2}\left(\alpha_{\mathrm{NHO}}\right) \tag{13.11}
$$

$$
\mathrm{E}_{\mathrm{HB2}} = \sum_{\substack{\forall\{i,j\} \in \\ \mathrm{ACC–H} \\ \forall\{m,n\} \in \\ \mathrm{ACC–H}}} w_{ij,mn}^{\mathrm{HB2}}\, \varepsilon_{ij,mn}^{\mathrm{HB2}} \exp\left(-\frac{\left(r_{ij} - r_{\mathrm{HB}}^{0}\right)}{2} \right) \exp\left(-\frac{\left(r_{mn} - r_{\mathrm{HB}}^{0}\right)}{2} \right) \tag{13.12}
$$

where E_{SB} is a sum over all pairs (ij) of negatively (Glu and Asp) and positively (Arg and

Lys) charged residues that can form a salt-bridge at neutral pH, E_{HB} is a sum over all pairs of acceptor and hydrogen linked to a donor that can form a hydrogen bond, and E_{HB2} is a sum over all pairs of hydrogen bonds satisfying the specific topological feature of an alpha-helix or a beta-sheet. E_{SB} is described by a repulsion-dispersion potential with a desolvation barrier (U_{DB}) similarly to C_α-C_α interactions. For simplicity, the strength of these interactions is controlled by a single weight (w^{SB}) for any type of salt-bridge in aaOPEP. E_{HB} is described by the same functional form as in OPEP [311] and it is non-zero only when the angle (α_{NHO}) is greater than 90°. The strength of a hydrogen bond is controlled by a weight (w_{ij}^{HB}) that depends on wether it is a main-chain/main-chain, main-chain/side-chain or side-chain/side-chain hydrogen bond. E_{HB2} is a sum over all pairs of main-chain/main-chain H-bonds (donor : don, acceptor : acc) satisfying (i) $\text{don}_1 = \text{acc}_1 + 4$, $\text{don}_2 = \text{acc}_2 + 4$ and $\text{acc}_2 = \text{acc}_1 + 1$ for alpha-helices, (ii) $\text{don}_1 = \text{acc}_2$ and $\text{don}_2 = \text{acc}_1 + 2$ for parallel β-sheets, and (iii) $\text{don}_1 = \text{acc}_2$ and $\text{don}_2 = \text{acc}_1$ or $\text{don}_2 = \text{don}_1 + 2$ et $\text{acc}_2 = \text{acc}_1 + 2$ for anti-parallel β-sheets. For simplicity, we only optimize one weight for beta-sheets, but we scale this weight by 0.25 for those that are parallel and by 0.75 for those that are anti-parallel to be representative of their respective population in the Protein Data Bank.

Overall, aaOPEP has 217 weights : 1 weight for each SC–SC pair type (w_{ij}^{SC-SC}, $21 \times 20/2 = 210$ weights in total), 1 weight for the C_α–C_α interaction ($w^{C\alpha-C\alpha}$), 1 weight for the salt-bridge interaction (w^{SB}), 1 weight for the main-chain/main-chain hydrogen bonded interaction (w^{mmHB}), 1 weight for the side-chain/side-chain hydrogen bonded interaction (w^{ssHB}), 1 weight for the side-chain/main-chain hydrogen bonded interaction (w^{smHB}), and 2 weights for the cooperativity between main-chain hydrogen bonds (w_{HB2}^{α} and w_{HB2}^{β}).

Identification of the peptide targets. The peptide targets were picked from the Protein Data Bank [43] using the following criteria : (i) their native state is a monomeric peptide solved by NMR in a solution of pH 5.5 to 7.1, (ii) they have less than 70 amino acids, (iii) they do not have non standard amino acids including termini cappings, (iv) they do not have disulfide bonds, (v) they are not membrane or amyloid peptides, and (vi) they are not solved with their ligand binding partner in the case of ligand- or ion-binding peptides. Moreover, the kClust algorithm is used to keep only a subset of peptides with less than 30% sequence identity [183].

We identify 2186 peptides with less than 50 amino acids before any filtering. After filtering, 131 peptides satisfy all our criteria and, after kClust, 90 peptides are left with less than 30% sequence identity. We also identify 1394 peptides having between 51 and 70 amino acids before

any filtering. After filtering and kClust, 206 peptides with less than 30% sequence identity remain. Of these 206 peptides, manual inspection and SCOPe identity [11, 12, 66, 160] revealed that there is 99 unique folds.

The training set consists of 50 peptides with more than 50 amino acids as they have more complex folds and populate a greater variety of side-chain/side-chain interaction types. To determine these 50 peptides from the 99 peptides, we use a Monte Carlo procedure that finds the optimal subset of peptides such that the population of side-chain/side-chain contact types is maximized.

Generation of the decoys. We employ different methods to generate a variety of decoys that are then energy minimized in the appropriate energy representation (sOPEP or aaOPEP). First, we use PEP-FOLD to generate de novo decoys in the sOPEP representation from the amino acid sequence without any constrains. Second, we use three kinds of molecular dynamics (MD) simulations at 300K starting from the native state of each peptide and the models generated using PEP-FOLD : (i) MD without restraint, (ii) MD with restraints on the ϕ and ψ dihedral angles of the regions that are participating in secondary structure motifs, and (iii) MD with restraints on the ϕ and ψ dihedral angles of the amino acids in the rigid core of the structure. The restraint applied on each dihedral angle is a quadratic–flat–quadratic potential with a flat region of $\pm 0.5^{\circ}$ around the initial angle value and a recoil constant of 250 kcal/mol/deg^2 for the quadratic part. When restraining the amino acids involved in secondary structure motifs, every possible combinations of restrains are considered. For example, there is seven independent simulations for restraining a three-helix bundle : restraining helices 1, 2, 3, 12, 13, 23 or 123. The rigid core of each peptide is defined as the residues that have a RMSf less than 1.5 Å as determined from its NMR models.

The decoys for aaOPEP (all-atom) are directly builded from the decoys generated for sOPEP (coarse-grained). The side-chain atoms are reconstructed using the SCWRL4 protocol [252]. The all-atom decoys are then energy minimized in the aaOPEP representation to remove any atomic clashes.

Classification of the decoys. The decoys are classified according to their similarity to the native state that is evaluated using the BCscore metric [178]. The BCscore is defined as the normalized sum of the signed volume of the parallelepipeds formed by all possible C_α triplets and the geometric center of the peptide. It runs from -1.0 to 1.0 where the latter value means that the two structures are the same and where the former value means that the two structures

are a perfect mirror image. We previously found [177, 430] that this scoring function is better at identifying peptides with similar folds compared to the RMSD and the TM-score [533].

We classify the decoys into three categories according to their BCscore with respect to the rigid core of the native state : the native-like (between 0.9 to 1.0), the near-native (between 0.7 to 0.9) and the non-native (smaller than 0.7). These intervals and the classification into three categories are inspired from previous works [311, 430]. Once the decoys are classified, we purge them to remove the ones that are too similar. A different BCscore threshold is used depending on the category : 0.98 for the native-like, 0.95 for the near-native and 0.90 for the non-native. When similar decoys are found, only the lowest energy one is kept. Finally, visual inspection of the BCscore distribution for the decoys of each peptide indicates that they well cover the BCscore spectrum, particularly between 0 and 1. Decoys with outlying energies are discarded.

Optimization protocol. The 217 weights of sOPEP and the 217 weights of aaOPEP are optimized using a genetic algorithm in order to stay consistent with the protocol used to optimize OPEP [311] and the previous sOPEP version [310]. The genetic algorithm optimizes the weights by evolving them through mutations and recombinations in order to best satisfy a set of pre-determined constrains. In our case, the constrains are that the energy of each non-native decoy should be higher than the energies of all near-native and native-like decoys, while the energy of each near-native decoy should be higher than the energies of all native-like decoys. The main differences with the protocol previously used are that (i) the lowest energy class consists of more than one structure i.e. the native-like states versus the native state before, (ii) the scoring function of the genetic algorithm is normalized such that the native-like/near-native, native-like/non-native and near-native/non-native constraints have equal weighting and (iii) the scoring function of the genetic algorithm is also normalized such that all peptide targets are equal weighting independently of the number of decoys they have.

The genetic algorithm that we use produces standard mutations with a rate of 20% that decreases by 0.25% every 10 iterations and single cross-over combination with a rate of 14%. We use a population of 100 chromosomes. We assume that the convergence is reached when the total number of satisfied constrains has not changed for 100 iterations. We repeat this procedure 10 times with different random seeds to ensure the convergence of the weights. At the end of each optimization, 50 Monte Carlo simulated annealing simulations of 50 000 steps are performed on the best weight set to confirm the convergence. The initial weights are inspired from OPEP [311].

The optimization protocol is separated into three sequential phases : (i) optimization of the side-chain/side-chain weights while keeping the other weights fixed, (ii) optimization of the H-bond weights while keeping the other weights fixed and (iii) optimization of the remaining weights while keeping the other weights fixed. This is necessary to obtain a better convergence of our results.

Validation protocol. For sOPEP, we use the validation set to confirm that it is able to better discriminate the native-like, near-native and non-native decoys. More precisely, we want more constrains to be satisfied using the optimized weights compared to the previous weights. For aaOPEP, we use the validation set to confirm that the decoys that are near from the native state can be refined. More precisely, we want to improve the structural quality of the decoys with an initial BCscore greater than 0.7 using restrained molecular dynamics simulations with aaOPEP. In these simulations, only the dihedral angles ϕ and ψ of the amino acids participating to secondary structures are restrained using a quadratic–flat–quadratic potential with a flat part of $\pm 0.5^{\circ}$ around the initial angle value and a recoil constant of 100 kcal/mol/deg^2 for the quadratic part.

Simulation software for sOPEP and aaOPEP. The simulation software used to prepare the decoys and to refine the structural prediction of PEP-FOLD is described in Annexe VI. This new software has been developed in our research group.

13.4 Perspective

At the moment, we have built the ensemble of decoys and obtained the first set of optimized weights for sOPEP and aaOPEP. Tests show that the optimized parameters are not ideal. Careful investigation and fine tuning of the different steps in the optimization protocol will be pursue by other students in the groups of professors Normand Mousseau and Pierre Tufféry. Below, we describe the expectations of the current project.

The first optimization of sOPEP was done using only 13 proteins in the training set. Not all side-chain/side-chain interaction types were then populated in the native states. Moreover, there were a lot more native/non-native constrains than constrains involving the native state as it was only a single structure. Both of these limitations could have biased the optimization procedure of the genetic algorithm. Our new approach reduces these limitations by (i) using a more extensive training set counting thousands of decoys for a total of more than 50 peptides, (ii) considering many native-like states (instead of just one) to account for the physiological flexibi-

lity of the protein, (iii) renormalizing the scoring function of the genetic algorithm such that the native-like/near-native, native-like/non-native and near-native/non-native constraints have equal weighting and (iv) renormalizing the scoring function of the genetic algorithm such that the peptide targets have equal weighting independently of the number of decoys they have. With our new protocol, we expect the new training set to produce an improved version of sOPEP that is more discriminative.

We also want to test the effect of the following modifications to sOPEP : (i) a less repulsive side-chain/side-chain interaction by reducing the R_{ij}^0 values (see Equation 3.4), (ii) a side-chain/side-chain interaction function more rapidly converging to zero – using a switch function – as the distance between the two side-chains increases, and (iii) a C_α–C_α interaction function that has the same form as the new side-chain/side-chain interaction. These modifications have been implemented and are currently being tested. With all that, we hope that a better sOPEP scoring function will increase the efficiency of the PEP-FOLD web server for de novo structure prediction [309, 310, 430, 461, 462] and of the PEP-SiteFinder web server for finding peptide-protein binding poses [407].

For the moment, the main goal of aaOPEP is to be used in a restrained molecular dynamics protocol to refine the structure prediction of both PEP-FOLD and PEP-SiteFinder web servers. In the future, we will continue to develop aaOPEP so that it might be used to investigate protein folding, function and aggregation. Stay tuned !

13.5 Author contributions

S.C. and P.T. designed the research ; S.C. designed and implemented aaOPEP ; S.C., V.B. and P.T. performed the research ; S.C. contributed analytic tools ; S.C., V.B., P.T., P.D. and N.M. analyzed the data ; S.C. and N.M. wrote the manuscript.

13.6 Acknowledgments

This work was funded in part by the Canada Research Chairs program, the Fonds de recherche du Québec – Nature et technologies (FRQNT), the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Fonds de recherche du Québec – Santé (FRQS). S.C. is grateful to Calcul Québec and the FRQNT for funding his internship in P.T.'s laboratory.

CHAPITRE 14

CONCLUSION

Les protéines sont les nanomachines moléculaires responsables du fonctionnement de nos cellules et sont ainsi au coeur de la vie. Elles se replient en structures tridimensionnelles qui définissent des fonctions dictées, bien sûr, par les lois de la physique. Dans le cas des protéines amyloïdes, elles s'agrègent en fibres amyloïdes très ordonnées et en structures amorphes associées à certaines maladies neurodégénératives telles que les maladies d'Alzheimer et de Huntington. Une compréhension des mécanismes de repliement et d'agrégation des protéines est nécessaire pour rationaliser leurs rôles dans la cellule. C'est ici que diverses méthodes de simulations jumelées à des modèles physiques entrent en jeu afin de décrire le repliement et l'agrégation des protéines avec une résolution atomique.

Le coeur de cette thèse s'articulait autour de deux axes de recherche principaux. D'une part, nous avons appliqué des méthodologies de simulation déjà existantes pour caractériser le repliement et l'agrégation des protéines amyloïdes bêta-amyloïde (Aβ) et huntingtine. Nous avons décrit, au niveau atomique, plusieurs des processus d'agrégation de ces deux protéines amyloïdes (Figure 14.1). D'autre part, nous avons développé un nouveau champ de force jumelé à un nouveau logiciel de dynamique moléculaire pour étudier le repliement et l'agrégation des protéines en général. Nous présentons ici les conclusions principales et les perspectives de recherche pour chacun de nos travaux.

14.1 Les premières étapes d'auto-assemblage de la protéine Aβ

La protéine bêta-amyloïde (Aβ), bien que ses rôles physiologiques ne soient pas clairement connus, attire une attention considérable étant donné son association à la maladie d'Alzheimer [179, 417]. Cette protéine forme de petits agrégats (oligomères) neurotoxiques et des fibres amyloïdes par des mécanismes qui ne sont pas très bien compris au niveau atomique. Les méthodes expérimentales à haute résolution actuelles sont, en effet, difficilement applicables puisque les petits oligomères de Aβ adoptent des ensembles structurels très hétérogènes et qu'ils ont tendance à s'agréger très rapidement. Or, une meilleure compréhension des mécanismes d'agrégation est nécessaire autant d'un point de vue théorique, pour mieux comprendre

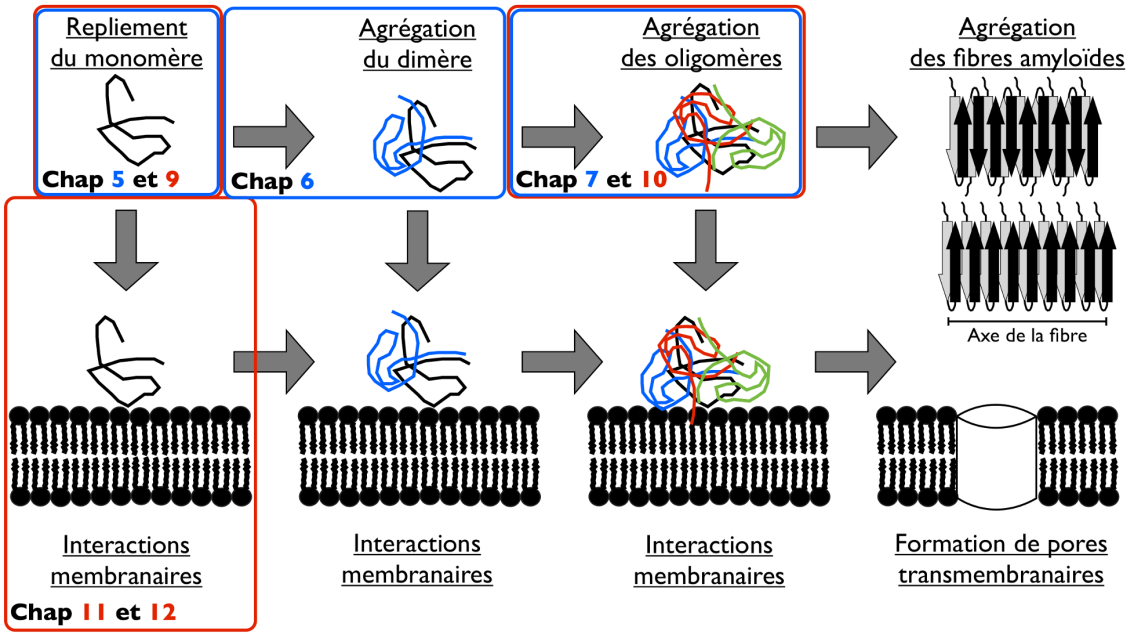


Figure 14.1 – Illustration des chemins d'agrégation caractérisés dans cette thèse pour les protéines amyloïdes bêta-amyloïde (Aβ) et huntingtine. Les systèmes étudiés à chacun des chapitres sont indiqués en bleu pour Aβ et en rouge pour huntingtine.

les processus de naissance et de croissance des fibres amyloïdes, que d'un point de vue pratique, pour le développement de composés thérapeutiques qui pourraient inhiber la formation des petits oligomères neurotoxiques.

Nous avons donc caractérisé, en solution aqueuse, le repliement du monomère et l'agrégation du dimère de trois formes physiologiques – Aβ_{40}, Aβ_{42} et Aβ_{40}(D23N) – de la protéine bêta-amyloïde (Aβ) à l'aide de simulations [108, 109]. Celles-ci permettent d'identifier, au niveau atomique, des différences structurelles significatives entre ces trois formes qui s'agrègent, dès la formation du dimère, par des mécanismes distincts. Le contexte et la méthodologie de nos simulations ont été expliqués en détail au Chapitre 4. Nous présentons maintenant les conclusions principales des articles sur le monomère et le dimère qui sont respectivement les Chapitres 5 et 6.

Le monomère des trois formes physiologiques de Aβ étudiées adopte un ensemble structurel principalement désordonné – peu de structures secondaires – en accord avec l'expérience. Les structures formées sont, au contraire, principalement stabilisées par des interactions non-polaires impliquant les régions CHC (acides aminés 17 à 21 dans la séquence de Aβ) et C-terminal

(acides aminés 29 à 40/42 dans la séquence de Aβ) pour les trois formes physiologiques. Malgré ces similitudes, des différences marquées existent néanmoins. La présence de deux acides aminés non-polaires (IA) supplémentaires au C-terminal de Aβ_{42} par rapport à Aβ_{40} possède des conséquences structurelles dès le monomère : Aβ_{42} a une propension plus grande à former des feuillets-β ainsi que des contacts non-polaires au CHC et au C-terminal. L'impact de la mutation Iowa Aβ_{40}(D23N) est moins drastique : la population de feuillets-β est légèrement plus élevée au C-terminal, formant des motifs tertiaires qui ne sont pas observés pour Aβ_{40}, et les contacts sont moins importants entre le N-terminal (acides aminés 1 à 16) et le *fibril loop* (acides aminés 22 à 28).

La formation du dimère est principalement alimentée par la séquestration des acides aminés non-polaires du CHC (acides aminés 17 à 21) et du C-terminal (acides aminés 29 à 40/42), ce qui est dû, entre autres, à la formation marquée de contacts intermoléculaires entre les régions CHC. L'ensemble structurel du dimère des trois formes physiologiques s'en trouve significativement modifié avec l'apparition d'une plus grande propension de feuillets-β au C-terminal et, pour Aβ_{42} seulement, au CHC. Les différences entre les formes physiologiques étudiées sont encore plus marquées dans le dimère en comparaison au monomère : Aβ_{42} est significativement plus structuré et il forme plus de contacts non-polaires que Aβ_{40}. Quant à lui, Aβ_{40}(D23N) est plus significativement structuré que Aβ_{40} au C-terminal.

Quels sont les impacts potentiels de ces observations sur l'agrégation de Aβ ? L'importance des régions CHC et C-terminal lors de l'agrégation de Aβ jusqu'à la structure de la fibre amyloïde a clairement été observée expérimentalement [44, 45, 55, 294, 469, 471]. Nos simulations montrent que ces régions forment, dès le monomère, des feuillets-β ainsi que la plupart des contacts entre les chaînes latérales des acides aminés. Nos simulations suggèrent aussi que le N-terminal, bien que n'étant pas structuré dans la fibre amyloïde, est nécessaire pour stabiliser le dimère en réduisant l'accessibilité au solvant des acides aminés non-polaires du CHC et du C-terminal. Ceci pourrait expliquer l'origine des chemins d'agrégation très différents qui sont observés expérimentalement en l'absence des premiers acides aminés du N-terminal (Aβ_{11-40}) [55].

L'impact des deux derniers acides aminés (IA) de Aβ_{42}, qui sont absents de Aβ_{40}, est significatif : augmentation de la fréquence et de la diversité des motifs feuillets-β et du nombre de contacts non-polaires. L'ensemble structurel du dimère Aβ_{42}, qui est déjà significativement

plus structuré que Aβ_{40}, pourrait favoriser une agrégation plus rapide de Aβ_{42} tel qu'observé expérimentalement. À l'opposé, Aβ_{40} devrait vraisemblablement subir des changements supplémentaires afin de mieux séquestrer ses acides aminés non-polaires. Ceci pourrait rationaliser l'origine de la forme globale très différente du tétramère (deux dimères) de ces deux formes physiologiques [45].

L'impact de la mutation Iowa Aβ_{40}(D23N) est moins évidente au niveau du monomère et du dimère, bien que l'expérience montre que cette forme physiologique s'agrège aussi beaucoup plus rapidement que Aβ_{40}. Par exemple, nous remarquons que le dimère de Aβ_{40}(D23N) n'a pas de feuillets-β au CHC, une région pourtant cruciale pour l'agrégation [284]. Par contre, il a une propension un peu plus grande à former au C-terminal, une autre région cruciale pour l'agrégation, des feuillets-β qui sont impliqués dans des motifs tertiaires plus variés. Par ailleurs, la mutation D23N est reconnue expérimentalement pour sa capacité à favoriser la formation de la fibre amyloïde par d'autres chemins d'agrégation ne dépendant pas de la formation du pont salin D23–K28, qui est un événement limitant dans la croissance de la fibre amyloïde de Aβ_{40} [416]. Or, le rôle de cet acide aminé n'est pas significatif au niveau du monomère et du dimère : la propension de D23 à interagir avec K28 est importante, mais l'interaction est plutôt faible étant donné qu'ils sont très accessibles au solvant. En somme, l'impact de cette mutation semble se faire sentir plus tard dans le processus d'agrégation.

À partir de ces observations, quels sont les mécanismes similaires se produisant durant l'agrégation de ces trois formes physiologiques de Aβ qui pourraient être ciblés par des composés thérapeutiques ? Trois modes d'action semblent privilégiés actuellement pour la création de nouveaux composés médicinaux : (i) ils interagissent avec les fibres amyloïdes pour diminuer leur aptitude à se fragmenter en oligomères, (ii) ils réduisent la stabilité des oligomères en accélérant la formation de fibres amyloïdes, ou (iii) ils interagissent directement avec les petits oligomères pour empêcher la formation des structures neurotoxiques. Les résultats de nos simulation offrent des pistes de réflexion concernant la dernière de ces trois catégories. Ceux-ci montrent que, durant le processus de formation du dimère, les trois formes physiologiques (i) adoptent une région C-terminal plus étendue avec une population de feuillets-β plus élevée, (ii) forment plusieurs contacts intermoléculaires dans la région CHC, (iii) possèdent moins de contacts intramoléculaires C-terminal/CHC et (iv) possèdent plus de contacts intramoléculaires CHC/C-terminal et intermoléculaires C-terminal/C-terminal. Ces tendances sont aussi obser-

vées dans d'autres simulations sur l'oligomérisation de Aβ_{40} et de Aβ_{42} ainsi que pour d'autres formes physiologiques (Chapitre 7) [340]. Pour empêcher ou modifier la formation de l'ensemble structurel du dimère, nos résultats suggèrent que les composés développés devraient interagir principalement avec les acides aminés du CHC (acides aminés 17 à 21, LVFFA) et du C-terminal (acides aminés 29 à 40/42, AIIGLMVGGVV/IA) pour empêcher ou modifier la formation des structures menant à la formation du dimère. Récemment, il a été observé que les polyphénols semblent être des composés prometteurs à cet égard selon des résultats provenant autant d'expériences que de simulations [340].

La méthode OPEP-REMD offre des perspectives intéressantes pour l'étude de l'agrégation de plus gros oligomères de Aβ grâce à son coût computationnel peu élevé. En effet, l'ensemble structurel de ces agrégats ne peut être échantillonné adéquatement dans des temps de simulations raisonnables en utilisant des modèles tout-atome avec une représentation explicite du solvant. Les structures gros-grains ainsi obtenues avec OPEP-REMD peuvent ensuite être converties en tout-atome afin de servir de point de départ dans des simulations utilisant une représentation plus complète. Une approche multi-échelle comme celle-ci a déjà été appliquée à l'étude de la formation de protofibres par le petit peptide amyloïde GNNQQNY [337]. Les structures tout-atome ainsi obtenues pourraient ensuite être mises en présence de composés thérapeutiques afin de caractériser leur aptitude à perturber la stabilité d'oligomères de Aβ plus gros que le dimère, ce qui n'a pas encore été réalisé. Par ailleurs, le protocole OPEP-REMD pourrait aussi très bien s'appliquer à l'étude de la croissance des fibres amyloïdes, ce qui permettrait d'étudier les deux autres modes d'actions des composés thérapeutiques qui agissent au niveau de la fibre. Le nouveau code de dynamique moléculaire que nous avons développé (opep_sim) ouvre cette perspective en permettant la simulation de plus gros systèmes moléculaires avec OPEP ainsi que l'application de restreintes variées durant la simulation (Annexe VI).

14.2 L'ensemble structurel et les interactions membranaires de huntingtine

La protéine huntingtine possède plusieurs rôles physiologiques dans la cellule [540]. Son N-terminal est étroitement associé à la régulation de ses fonctions et contient quatre segments : Htt17 qui est amphiphile, Q_N qui est amyloïdogénique, P_{11} qui contient onze prolines consécutives et un dernier segment qui est riche en prolines. Le segment Htt17 contrôle la localisation de huntingtine dans la cellule en ancrant la protéine sur des membranes lipidiques. Le segment

Q_N, quant à lui, est composé d'un nombre de répétitions, qui varie d'une personne à l'autre, de l'acide aminé glutamine. Lorsqu'il est composé de plus de 36 glutamines, il est responsable du mauvais repliement et de la localisation incorrecte de huntingtine ainsi que de son agrégation en fibres amyloïdes et en oligomères neurotoxiques qui sont associés à la maladie de Huntington [358, 539]. La structure du N-terminal de la protéine huntingtine est inconnue tant au niveau du monomère que des oligomères, car il adopte des structures hétérogènes principalement désordonnées qui ont tendance à s'agréger facilement. Ceci rend donc les méthodes expérimentales à haute résolution difficilement applicables. Or, une compréhension de ces structures est nécessaire pour mieux comprendre les rôles physiologiques et pathologiques de huntingtine à travers son ensemble structurel en solution et ses interactions membranaires.

Nous avons donc caractérisé le repliement du monomère de fragments du N-terminal de huntingtine en solution aqueuse ainsi qu'en présence d'une membrane de phospholipides. Nos simulations permettent d'identifier la présence de motifs structuraux qui pourraient favoriser son agrégation et ses interactions membranaires ainsi que leur dépendance envers ses régions : Htt17, Htt17Q_N ou Htt17$Q_N P_{11}$. Le contexte et la méthodologie de nos simulations ont été expliqués en détail au Chapitre 8. Nous présentons maintenant les conclusions principales des articles sur le monomère de Htt17, Htt17Q_{17} et Htt17$Q_{17}P_{11}$ (Chapitre 9), sur les structures nanotubulaires de Q_{30}, Q_{40}, Htt17Q_{30} et Htt17Q_{40} (Chapitre 10) ainsi que les interactions membranaires de Htt17 et Htt17Q_{20} (Chapitres 11 et 12).

Au niveau du monomère [52], nos résultats sur Htt17 concordent avec les données expérimentales tout en les complétant grâce une caractérisation de l'ensemble structurel accessible à la protéine. Htt17 adopte un assemble assez hétérogène de structures désordonnées en partie hélicoïdales. En particulier, les premiers acides aminés de sa séquence ont une plus grande propension à former des hélices-α. Par ailleurs, les acides aminés non-polaires sont accessibles au solvant. Ces deux caractéristiques pourraient favoriser l'agrégation de Htt17 ainsi que ses interactions membranaires. En effet, l'expérience suggère que Htt17 forme des tétramères d'hélices-α en solution, tandis que l'expérience et les simulations montrent que le monomère de Htt17 adopte intégralement une hélice-α lorsqu'il est inséré dans une membrane. L'ajout du segment Q_N provoque des changements significatifs à la structure de Htt17, dont la propension d'hélice-α qui est décalée aux derniers acides aminés de Htt17. Par ailleurs, les acides aminés non-polaires de Htt17 deviennent moins accessibles au solvant. L'ajout du segment P_{11} a un impact encore

plus significatif : Htt17 devient beaucoup plus structuré et ses acides aminés non-polaires plus accessibles au solvant. Ces observations suggèrent que les premières étapes d'agrégation de Htt17, Htt17Q_{17} et Htt17$Q_{17}P_{11}$ seront différentes. Maintenant que nous avons caractérisé le repliement du monomère, la prochaine étape sera de regarder la formation du tétramère avec un protocole de simulation similaire (HREXMetaD) afin d'évaluer le modèle expérimental du professeur Ronald Wetzel qui suggère que Htt17 s'assemble en tétramère, ce qui favorise la formation de feuillets-β dans Q_N en augmentant la concentration locale de peptides [501].

Au niveau des oligomères [110], nos résultats montrent que la structure nanotubulaire est plus stable pour Q_{40} que pour Q_{30}. De plus, la croissance de ces structures se fait à partir d'une sous-unité dimère et elle s'opère par la formation de ponts-H. Finalement, ces nanotubes sont plus stables lorsque la région Htt17 est considérée (Htt17Q_{30} et Htt17Q_{40}). Il s'agit d'une observation importante, car Htt17 est reconnu pour accélérer l'agrégation de Q_N. Or, le modèle le plus répandu suggère que Htt17 s'agrège en tétramère, ce qui amène les segments Q_N à proximité y favorisant l'apparition de feuillets-β [501]. Nos résultats suggèrent un nouveau mécanisme, à tout le moins pour les structures nanotubulaires simulées, par lequel Htt17 pourrait favoriser l'agrégation en stabilisant la présence de feuillets-β dans Q_N. Il serait intéressant de confirmer que la même observation s'applique aussi à la structure de la fibre amyloïde de huntingtine qui a été déterminée par RMN [414]. Par ailleurs, bien que la structure nanotubulaire ne soit pas complètement acceptée par la communauté scientifique [400], des expériences récentes par la diffusion de neutrons aux petits angles de neutrons ont permis l'observation de telles structures pour Htt17$Q_{42}P_{10}$, mais pas pour Htt17$Q_{22}P_{10}$ [372], ce qui supporte nos observations.

Au niveau des interactions membranaires du N-terminal de huntingtine [106, 111], nos résultats suggèrent que l'insertion complète de Htt17 et Htt17Q_N dans une membrane de phospholipides est favorisée par la présence d'hélices-α antérieure à leur fixation sur la membrane. De plus, le segment Q_N semble aider l'insertion du segment Htt17 en interagissant – formation de ponts-H – avec la tête polaire des phospholipides. L'ajout de Q_N ne semble pas modifier significativement l'orientation et la structure de Htt17 lorsqu'il est inséré dans la membrane. Nous avons ensuite effectué un ensemble important de simulations sur la région Htt17 afin de compléter, en collaboration, les résultats du groupe d'expérimentateurs du professeur Burkhard Bechinger [317–319]. Nos résultats suggèrent principalement que tous les acides aminés de Htt17 sont structurés sur une membrane de phospholipides, contrairement à ce que laisse sug-

gérer le modèle RMN obtenu en présence de micelles pour lequel seuls les acides aminés 6 à 17 forment une hélice-α. En effet, nous avons observé que le modèle RMN se replie complètement en hélice-α lorsqu'il est inséré dans une membrane de phospholipides. Nos simulations révèlent aussi l'accessibilité au solvant et la profondeur d'insertion dans la membrane de tous les acides aminés de la séquence de Htt17. Ceci complète les données expérimentales de nos collaborateurs qui ne sont disponibles que pour quelques acides aminés. De plus, nos simulations ont permis d'identifier les perturbations de la membrane ainsi que les acides aminés qui interagissent avec la membrane en terme de ponts-H, de ponts salins et d'interactions non-polaires, ce qui n'était pas encore connu.

Trois avenues de recherche sont envisageables sur les interactions huntingtine–membrane, un système peu étudié *in silico*. Premièrement, l'importance relative des interactions peptides–membrane pourrait être évaluée par des protocoles *in silico* de mutagenèse à l'alanine [360]. En partant des orientations principales de Htt17 sur une membrane observées dans nos simulations, nous pourrions appliquer ces protocoles et évaluer le coût en énergie libre de la mutation. Ceci nous permettrait d'ordonner les acides aminés en fonction de leur importance. Deuxièmement, les orientations de Htt17 dans nos simulations et celles observées expérimentalement suggèrent que Htt17 pourrait former un dimère sur une membrane de phospholipides. Afin d'évaluer cela, nous avions effectué un ensemble assez important de dynamiques moléculaires qui n'ont pas convergées adéquatement étant donné la complexité de ce système. La méthode HREXMetaD nous offre la possibilité de revenir à ce système pour caractériser la stabilité et la formation du dimère de Htt17 sur une membrane de phospholipides. Troisièmement, les caractères amphiphile et chargé de Htt17, similaires aux peptides anti-microbiens [194, 505], suggèrent qu'il pourrait former des pores transmembranaires dont la stabilité pourrait être influencée par le segment amyloïdogénique Q_N dû à ces interactions avec la tête des phospholipides. Plusieurs protéines amyloïdes, comme Aβ, forment effectivement des pores dans les membranes qui perturberaient l'équilibre physiologique des cellules [263, 264].

14.3 Les modèles sOPEP et aaOPEP

Le second volet de cette thèse a une portée future et des acquis significatifs, malgré la place moins importante qu'occupent ces travaux dans la présente thèse (Chapitre 13 et Annexe VI) [105, 107].

Nous avons développé un protocole de paramétrisation et un ensemble de leurres structurels pouvant être utilisés pour optimiser ou ré-optimiser n'importe quel modèle gros-grain [107]. Ce protocole sera, dans un premier temps, utilisé pour ré-optimiser le potentiel sOPEP, car il ne discrimine pas assez les structures natives des structures non-native, particulièrement pour les peptides de plus de 50 acides aminés. Dans un deuxième temps, le même protocole sera appliqué pour optimiser un nouveau modèle simplifié que nous avons développé pour les protéines. Ce modèle est une extension dans le domaine du tout-atome des principes fondateurs du modèle gros-grain OPEP qui est largement utilisé pour étudier le repliement et l'agrégation des protéines. L'utilité première du modèle tout-atome OPEP (aaOPEP) est de raffiner les prédictions *de novo* par PEP-FOLD de la structure de protéines directement à partir de leur séquence d'acides aminés. PEP-FOLD se compare déjà très avantageusement aux autres méthodes de prédiction *de novo* pour les peptides de moins de 50 acides aminés, mais il ne possède pas d'étape de raffinement contrairement à ces autres méthodes. L'ajout d'une étape de raffinement augmentera encore plus sa précision, particulièrement pour les peptides de 50 à 70 acides aminés. Les divers outils pour le projet d'optimisation des modèles sOPEP et aaOPEP ont été développés pour la procédure en entier. Nous sommes rendus à peaufiner les diverses étapes afin d'améliorer le résultat final des optimisations. Nous allons aussi évaluer de façon critique le choix des fonctionnelles pour aaOPEP par rapport à son habilité à discriminer les structures natives et non-natives. Par la suite, nous envisageons aussi d'ajuster les paramètres de aaOPEP pour des simulations à température constante.

En parallèle à nos travaux sur aaOPEP, nous avons développé un programme de dynamique moléculaire, appelé opep_sim, spécifiquement conçu pour la famille de potentiel OPEP – OPEPv3, OPEPv4, OPEPv5, sOPEP et aaOPEP – qui sera entre autres utilisé pour le raffinement des prédictions de PEP-FOLD [105]. Ce programme a été conçu pour être très facilement adaptable pour les futures versions de OPEP, sOPEP et aaOPEP. Il sera diffusé librement à ceux qui veulent l'utiliser, augmentant d'autant ses retombées. Les prochaines étapes de développement de opep_sim sont l'ajout de la parallélisation du calcul des forces – déjà planifiée dans l'élaboration du code – et de techniques d'échantillonnage avancées telles que la dynamique moléculaire d'échanges de répliques.

BIBLIOGRAPHIE

[1] J. L. F. Abascal et C. Vega. A general purpose model for the condensed phases of water : Tip4p/2005. *J. Chem. Phys.*, 123:234505, 2005.

[2] S. A. Adcock et J. A. McCammon. Molecular dynamics : Survey of methods for simulating the activity of proteins. *Chem. Rev.*, 106(106):1589–1615, 2006.

[3] M. Ahmed, J. Davis, D. Aucoin, T. Sato, S. Ahuja, S. Aimoto et al. Structural conversion of neurotoxic amyloid-beta(1-42) oligomers to fibrils. *Nat. Struct. Mol. Biol.*, 17(5):561–567, 2010.

[4] C. T. Aiken, J. S. Steffan, C. M. Guerrero, H. Khashwji, T. Lukacsovich, D. Simmons, J. M. Purcell, K. Menhaji, Y.-Z. Zhu, K. Green, F. LaFerla, L. Huang, L. M. Thompson et J. L. Marsh. Phosphorylation of threonine 3 - implications for huntingtin aggregation and neurotoxicity. *J. Biol. Chem.*, 284(43):29427–29436, 2009.

[5] C. Aisenbrey et B. Bechinger. Tilt and rotational pitch angle of membrane-inserted polypeptides from combined ^{15}n and ^{2}h solid-state nmr spectroscopy. *Biochemistry*, 43:10502–10512, 2004.

[6] D. M. J. Alexander, F. Michael et L. B. I. Charles. Extending the treatment of backbone energetics in protein force fields : limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.*, 25:1400–1415, 2004.

[7] W. J. Allen, J. A. Lemkul et D. R. Bevan. Gridmat-md : A grid-based membrane analysis tool for use with molecular dynamics. *J. Comput. Chem.*, 30(12):1952–1958, 2009.

[8] P. Anand, F. S. Nandel et U. H. E. Hansmann. The alzheimer's β amyloid ($a\beta_{1-39}$) dimer in an implicit solvent. *J. Chem. Phys.*, 129(19):195102, 2008.

[9] H. C. Andersen. Rattle : A "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.*, 52(1):24–34, 1983.

[10] M. A. Andrade et P. Bork. Heat repeats in the huntington's disease protein. *Nat. Genet.*, 11(2):115–116, 1995.

[11] A. Andreeva, D. Howorth, S. E. Brenner, T. J. P. Hubbard, C. Chothia et A. G. Murzin. Scop database in 2004 : refinements integrate structure and sequence family data. *Nucl. Acids Res.*, 32:D226–D229, 2004.

[12] A. Andreeva, D. Howorth, J.-M. Chandonia, S. E. Brenner, T. J. P. Hubbard, C. Chothia et A. G. Murzin. Data growth and its impact on the scop database ; new developments. *Nucl. Acids Res.*, 36:D419–D425, 2008.

[13] J. R. Arndt, S. G. Kondalaji, M. M. Maurer, A. Parker, J. Legleiter et S. J. Valentine. Huntingtin n-terminal monomeric and multimeric structures destabilized by covalent modification of heteroatomic residues. *Biochemistry*, 54(28):4285–4296, 2015.

[14] R. S. Atwal, C. R. Desmond, N. Caron, T. Maiuri, J. Xia, S. Sipione et R. Truant. Kinase inhibitors modulate huntingtin cell localization and toxicity. *Nat. Chem. Biol.*, 7:453–460, 2011.

[15] R. S. Atwal et R. Truant. A stress sensitive er membrane-association domain in huntingtin protein defines a potential role for huntingtin in the regulation of autophagy. *Autophagy*, 4(1):91–93, 2008.

[16] R. S. Atwal, J. Xia, D. Pinchev, J. Taylor, R. M. Epand et R. Truant. Huntingtin has a membrane association signal that can modulate huntingtin aggregation, cellular entry and toxicity. *Hum. Mol. Genet.*, 16(21): 2600–2615, 2007.

[17] D. Baker et A. Sali. Protein structure prediction and structural genomics. *Science*, 294:93–96, 2001.

[18] A. Barducci, M. Bonomi et P. Derreumaux. Assessing the quality of the opep coarse-grained force field. *J. Chem. Theory Comput.*, 7:1928–1934, 2011.

[19] A. Barducci, M. Bonomi et M. Parrinello. Metadynamics. *WIREs Comput. Mol. Sci.*, 1:826–843, 2011.

[20] A. Barducci, G. Bussi et M. Parrinello. Well-tempered metadynamics : a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100:020603, 2008.

[21] D. J. Barlow et J. M. Thornton. Ion-pairs in proteins. *J. Mol. Biol.*, 168:867–885, 1983.

[22] A. Barth et C. Zscherp. What vibrations tell us about proteins. *Q. Rev. Biophys.*, 35(4):369–430, 2002.

[23] S. Barton, R. Jacak, S. D. Khare, F. Ding et N. V. Dokholyan. The length dependence of the polyq-mediated protein aggregation. *J. Biol. Chem.*, 282(35):25487–25492, 2007.

[24] B. Barz, O. O. Olubiyi et B. Strodel. Early amyloid β-protein aggregation precedes conformational change. *Chem. Commun.*, 50(40):5373–5375, 2014.

[25] B. Barz et B. Urbanc. Dimer formation enhances structural differences between amyloid beta-protein (1-40) and (1-42) : an explicit-sovlent molecular dynamics study. *PLoS ONE*, 7(4):e34345, 2012.

[26] G. P. Bates. The molecular genetics of huntington disease – a history. *Nat. Rev. Genet.*, 6:768–773, 2005.

[27] A. Baumketner, S. L. Bernstein, T. Wyttenbach, N. D. Lazo, D. B. Teplow, M. T. Bowers et J.-E. Shea. Structure of the 21-30 fragment of amyloid β-protein. *Protein Sci.*, 15:1239–1247, 2006.

[28] K. A. Beauchamp, Y.-S. Lin, R. Das et V. S. Pande. Are protein force fields getting better ? a systematic benchmark on 524 diverse nmr measurements. *J. Chem. Theory Comput.*, 8:1409–1414, 2012.

[29] A. L. Beberg, D. L. Ensign, G. Jayachandran, S. Khaliq et V. S. Pande. Folding@home : Lessons from eight years of volunteer distributed computing. Dans *IEEE international symposium on parallel & distributed processing.*, pages 1–8. IEEE, 2009.

[30] B. Bechinger. A dynamic view of peptides and proteins in membranes. *Cell. Mol. Life Sci.*, 65:3028–3039, 2008.

[31] B. Bechinger. Rationalizing the membrane interactions of actinic amphipathic antimicrobial peptides by their molecular shape. *Curr. Opin. Colloid Interface Sci.*, 14:349–355, 2009.

[32] B. Bechinger. Membrane association and pore formation by alpha-helical peptides. Dans G. Anderluh et J. Lakey, éditeurs, *Proteins : membrane binding and pore formation.*, pages 24–30. Landes bioscience and springer science+business media, 2010.

[33] B. Bechinger et C. Aisenbrey. The polymorphic nature of membrane-active peptides from biophysical and structural investigations. *Curr. Protein Peptide Sci.*, 13:602–610, 2012.

[34] B. Bechinger, C. Aisenbrey et P. Bertani. The alignment, structure and dynamics of membrane-associated polypeptides by solid-state nmr spectroscopy. *BBA-Biomembranes*, 1666(1):190–204, 2004.

[35] B. Bechinger et C. Resende, J. M. Aisenbrey. The structural and topological analysis of membrane-associated polypeptides by oriented solid-state nmr spectroscopy : Established concepts and novel developments. *Biophys. Chem.*, 153(2-3):115–125, 2011.

[36] B. Bechinger et C. Sizun. Alignment and structural analysis of membrane polypeptides by ^{15}n and ^{31}p solid-state nmr spectroscopy. *Concepts Magn. Reso. A*, 18(2):130–145, 2002.

[37] I. Benilova, E. Karran et B. De Strooper. The toxic aβ oligomer and alzheimer's disease : an emperor in need of clothes. *Nat. Neurosci.*, 15(3):349–357, 2012.

[38] C. L. Benn, C. Landles, H. Li, A. D. Strand, B. Woodman, K. Sathasivam, S. H. Li, S. Ghazia-Noori, E. Hockly, S. M. Faruque et al. Contribution of nuclear and extranuclear polyq to neurological phenotypes in mouse models of huntington's disease. *Hum. Mol. Genet.*, 14:3065–3078, 2005.

[39] H. J. C. Berendsen, J. R. Grigera et T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91(24):6269–6271, 1987.

[40] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola et J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, 1984.

[41] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren et J. Hermans. Interaction models for water in relation to protein hydration. Dans B. Pullman, éditeur, *Intermolecular forces*, pages 331–342. D. Reidel Publishing, 1981.

[42] H. J. C. Berendsen, D. van der Spoel et R. van Drunen. Gromacs : a message-passing parallel molecular dynamics implementation. *Comp. Phys. Commun.*, 91:43–56, 1995.

[43] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov et P. E. Bourne. The protein data bank. *Nucleic Acids Res.*, 28:235–242, 2000.

[44] S. L. Bernstein, T. Wyttenbach, A. Baumketner, J.-E. Shea, G. Bitan, D. B. Teplow et M. T. Bowers. Amyloid β-protein : Monomer structure and early aggregation states of aβ42 and its pro^{19} alloform. *J. Am. Chem. Soc.*, 127(7):2075–2084, 2005.

[45] S.L. Bernstein, N. F. Dupuis, N. D. Lazo, T. Wyttenbach, M. M. Condron, G. Bitan, D. B. Teplow, J.-E. Shea, B. T. Ruotolo, C. V. Robinson et M. T. Bowers. Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of alzheimer's disease. *Nat. Chem.*, 1(4):326–331, 2009.

[46] R. B. Best et G. Hummer. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B*, 113:9004–9015, 2009.

[47] R. B. Best et J. Mittal. Protein simulations with an optimized water model : Cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B*, 114:14916–14923, 2010.

[48] R. B. Best, W. Zheng et J. Mittal. Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.*, 10:5113–5124, 2014.

[49] M. R. Betancourt et D. Thirumalai. Pair potentials for protein folding : Choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Prot. Sci.*, 8:361–369, 1999.

[50] M. Betnel, N. V. Dokholyan et B. Urbanc. From disordered amyloid β-protein to soluble oligomers and protofibrils using discrete molecular dynamics. Dans P. Derreumaux, éditeur, *Alzheimer's diseases : insights into low molecular weight and cytotoxic aggregates from in vitro and compute experiments.*, pages 333–357. Imperial College Press, 2013.

[51] A. Bhattacharyya, A. K. Thakur, V. M. Chellgren, G. Thiagarajan, A. D. Williams, B. W. Chellgren, T. P. Creamer et R. Wetzel. Oligoproline effects on polyglutamine conformation of aggregation. *J. Mol. Biol.*, 355:524–535, 2006.

[52] V. Binette, S. Côté et N. Mousseau. Free-energy landscape of the amino-terminal fragment of huntingtin in aqueous solution. *Submitted*, 2015.

[53] G. Bitan, M. D. Kirkitadze, A. Lomakin, S. S. Vollers, G. B. Benedek et D. B. Teplow. Amyloid β-protein (aβ) assembly : Aβ40 and aβ42 oligomerize through distinct pathways. *P. Natl. Acad. Sci. USA*, 100(1): 330–335, 2003.

[54] G. Bitan, A. Lomakin et D. B. Teplow. Amyloid β-protein oligomerization : pre nucleation interactions revealed by photo-induced cross-linking of unmodified proteins. *J. Biol. Chem.*, 276(37):35176–35184, 2001.

[55] G. Bitan, S. S. Vollers et D. B. Teplow. Elucidation of primary structure elements controlling early amyloid β-protein oligomerization. *J. Biol. Chem.*, 278(37):34882–34889, 2003.

[56] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler et P. Gumbsch. Structural relaxation made simple. *Phys. Rev. Lett.*, 97:170201, 2006.

[57] D. Boal. *Mechanics of the cell.* Cambridge University Press, 2002.

[58] K. Boesze-Battaglia. Membrane rafts. Dans P. L. Yeagle, éditeur, *The structure of biological membranes.*, pages 309–327. CRC Press LLC, 2004.

[59] M. Bokvist, F. Lindstrom, A. Watts et G. Grobner. Two types of alzheimer's β-amyloid (1-40) peptide membrane interactions : aggregation preventing transmembrane anchoring versus accelerated surface fibril formation. *J. Mol. Biol.*, 335:1039–1049, 2004.

[60] W. Bolton et M. F. Perutz. Three dimensional fourier synthesis of horse deoxyhaemoglobin at 2.8å resolution. *Nature*, 228:551–552, 1970.

[61] M. Bonomi, A. Barducci et M. Parrinello. Reconstructing the equilibrium boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem*, 30:1615–1621, 2009.

[62] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, Broglia R. A. et M. Parrinello. Plumed : a portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.*, 180:1961–1972, 2009.

[63] K. J. Bowers, E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossaváry, M. A. Moraes, F. D. Sacerdoti, J. K. Salmon, Y. Shan et D. E. Shaw. Scalable algorithms for molecular dynamics simulations on commodity clusters. Dans *Proceedings of the ACM/IEEE conference on supercomputing.*, page 13. IEEE, 2006.

[64] C. Branden et J. Tooze. *Introduction to Protein Structure*. Garland Publishing Inc., 2e édition édition, 1998.

[65] B. Brendan. Power play. *Nature*, 451:240–243, 2008.

[66] S. E. Brenner, C. Chothia, T. J. P. Hubbard et A. G. Murzin. Understanding protein structure : using scop for fold interpretation. *Methods Enzymol.*, 37:635–643, 1996.

[67] B. R. Brooks, C. L. Brooks III, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York et M. Karplus. Charmm : The biomolecular simulation program. *J. Comp. Chem.*, 30:1545–1615, 2009.

[68] B. R. Brooks, R. E. Bruccoler, B. D. Olafson, D. J. States, S. Swaminathan et M. Karplus. Charmm : a program for macromolecular energy minimization and dynamics calculations. *J. Comput. Chem.*, 4:187–217, 1982.

[69] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan et M. Karplus. Charmm : A program for macromolecular energy minimization and dynamics calculations. *J. Comp. Chem.*, 4:187–217, 1983.

[70] S. K. Burger, P. W. Ayers et J. Schofield. Efficient parametrization of torsional terms for force fields. *J. Comput. Chem.*, 35:1438–1445, 2014.

[71] K. A. Burke, K. M. Hensal, C. S. Umbaugh, M. Chaibva et J. Legleiter. Huntingtin disrupts lipid bilayers in a polyq-length dependent manner. *Biochim. Biophys. Acta*, 1828:1953–1961, 2013.

[72] K. A. Burke, K. J. Kauffman, C. S. Umbaugh, S. L. Frey et J. Legleiter. The interaction of polyglutamine peptides with lipid membranes in regulated by flanking sequences associated with huntingtin. *J. Biol. Chem.*, 288(21):14993–15005, 2013.

[73] K. A. Burke, E. A. Yates et J. Legleiter. Amyloid-forming proteins alter the local mechanical properties of lipid membranes. *Biochemistry*, 52(5):808–817, 2013.

[74] G. Bussi. Hamiltonian replica exchange in gromacs : a flexible implementation. *Mol. Phys.*, 112(3-4): 379–384, 2014.

[75] G. Bussi, D. Donadio et M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126: 014101, 2007.

[76] C. Camilloni, D. Provasi, G. Tiana et R. A. Broglia. Exploring the protein g helix free-energy surface by solute tempering metadynamics. *Proteins*, 71(4):1647–1654, 2008.

[77] C. Camilloni, D. Schaal, K. Schweimer, S. Schwarzinger et A. De Simone. Energy landscape of the prion protein helix 1 probed by metadynamics and nmr. *Biophys. J.*, 102(1):158–167, 2012.

[78] D. A. Case, J. T. Berryman, R. M. Betz, D. S. Cerutti, T. E. Cheatham, III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K. M. Merz, G. Monard, P. Needham, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, R. Salomon-Ferrer, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, D. M. York et P. A. Kollman. Amber 2015. *University of California, San Francisco*, 2015.

[79] J. P. Caviston et E. L. F. Holzbaur. Huntingtin as an essential integrator of intracellular vesicular trafficking. *Trends Cell. Biol.*, 19(4):147–155, 2009.

[80] M. Cernescu, T. Stark, E. Kalden, C. Kurz, K. Leuner, T. Deller et al. Laser-induced liquid bead ion desorption mass spectrometry : an approach to precisely monitor the oligomerization of the beta-amyloid peptide. *Anal. Chem.*, 84(12):5276–5284, 2012.

[81] D. S. Cerutti, W. C. Swope, J. E. Rice et D. A. Case. ff14ipq : A self-consistent force field for condensed-phase simulations of proteins. *J. Chem. Theory Comput.*, 10:4515–4534, 2014.

[82] M. Chaibva, K. A. Burke et J. Legleiter. Curvature enhances binding and aggregation of huntingtin at lipid membranes. *Biochemistry*, 53:2355–2365, 2014.

[83] D. I. Chan, E. J. Prenner et H. J. Vogel. Tryptophan- and arginine-rich antimicrobial peptides : structures and mechanisms of action. *Biochim. Biophys. Acta*, 1758:1184–1202, 2006.

[84] H. S. Chan, Z. Zhang, S. Wallin et Z. Liu. Cooperativity, local-nonlocal coupling, and nonnative interactions : principles of protein folding from coarse-grained models. *Annu. Rev. Phys. Chem.*, 62:301–326, 2011.

[85] D. E. Chapman, J. K. Steck et P. S. Nerenberg. Optimizing protein-protein van der waals interactions for the amberff9x/ff12 force field. *J. Chem. Theory Comput.*, 10:273–281, 2014.

[86] Y. Chebaro, X. Dong, R. Laghaei, P. Derreumaux et N. Mousseau. Replica exchange molecular dynamics simulations of coarse-grained proteins in implicit solvent. *J. Phys. Chem. B*, 113:267–274, 2009.

[87] Y. Chebaro, P. Jiang, T. Zang, Y. Mu, P. H. Nguyen, N. Mousseau et al. Structures of aβ17-42 trimers in isolation and with five small-molecular drugs using a hierarchical computational procedure. *J. Chem. Phys. B*, 116(29):8412–8422, 2012.

[88] Y. Chebaro, N. Mousseau et P. Derreumaux. Structures and thermodynamics of alzheimer's amyloid-β aβ(16-35) monomer and dimer by replica exchange molecular dynamics simulations : implication for full-length aβ fibrillation. *J. Phys. Chem. B.*, 113(21):7668–7675, 2009.

[89] Y. Chebaro, S. Pasquali et P. Derreumaux. The coarse-grained opep force field for non-amyloid and amyloid proteins. *J. Phys. Chem. B*, 116:8741–8752, 2012.

[90] S. Chen, V. Berthelier, W. Yang et R. Wetzel. Polyglutamine aggregation behaviour in vitro supports a recruitment mechanism of cytotoxicity. *J. Mol. Biol.*, 311:173–182, 2001.

[91] S. Chen, F. A. Ferrone et R. Wetzel. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *P. Natl. Acad. Sci. USA*, 99(18):11884–11889, 2002.

[92] W. Chen, N. Mousseau et P. Derreumaux. The conformations of the amyloid-β (21-30) fragment can be described by three families in solution. *J. Chem. Phys*, 125(8):084911, 2006.

[93] Y.-R. Chen et C. G. Glabe. Distinct early folding and aggregation properties of alzheimer amyloid-β peptides aβ40 and aβ42. *J. Biol. Chem.*, 281(34):24414–24422, 2006.

[94] Y.-R. Chen, H.-B. Huang, C.-J. Lo, C.-C. Wang, C.-L. Su, H.-T. Liu, M.-S. Shiao, T.-H. Lin et Y.-C. Chen. Aβ_{40}(l17a/f19a) mutant diminishes the aggregation and neurotoxicity of aβ_{40}. *Biochem. Biophys. Res. Co.*, 405:91–95, 2011.

[95] M. S. Cheung, A. E. Garcia et J. N. Onuchic. Protein folding mediated by solvation : water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. USA*, 99: 685–690, 2002.

[96] J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok et K. A. Dill. Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. *J. Chem. Theory Comput.*, 3(1): 26–41, 2007.

[97] M. Chopra, A. S. Reddy, N. L. Abbott et J. J. de Pablo. Folding of polyglutamine chains. *J. Chem. Phys.*, 129:135102, 2008.

[98] E. A. Cino, W.-Y. Choy et M. Karttunen. Comparison of secondary structure formation using 10 different force fields in microsecond molecular dynamics simulations. *J. Chem. Theory Comput.*, 8:2725–2740, 2012.

[99] G. Cisbani et F. Cicchetti. An *in vitro* perspective on the molecular mechanisms underlying mutant huntingtin protein toxicity. *Cell Death Dis.*, 3:e382, 2012.

[100] M. Compiani et E. Capriotti. Computational and theoretical methods for protein folding. *Biochemistry*, 52: 8601–8624, 2013.

[101] S. H. Cong et S. Ham. Atomic-level investigations on the amyloid-beta dimerization process and its driving forces in water. *Phys. Chem. Chem. Phys.*, 14(5):1573–1575, 2012.

[102] A. Cordomí, G. Caltabiano et L. Pardo. Membrane protein simulations using amber force field and berger lipid parameters. *J. Chem. Theory Comput.*, 8:948–958, 2012.

[103] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Jr. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell et P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.*, 117(19):5179–5197, 1995.

[104] J. Cornett, F. Cao, C.-E. Wang, C. A. Ross, G. P. Bates, S.-H. Li et X.-J. Li. Polyglutamine expansion of huntingtin impairs its nuclear export. *Nat. Genet.*, 37(2):198–204, 2005.

[105] S. Côté, V. Binette, L. Jutra-Dubé et N. Mousseau. A molecular dynamics program for the family of force fields opep. *In preparation.*, 2016.

[106] S. Côté, V. Binette, E. S. Salnikov, B. Bechinger et N. Mousseau. Probing the huntingtin 1-17 membrane anchor on a phospholipid bilayer by using all-atom simulations. *Biophys. J.*, 108:1187–1198, 2015.

[107] S. Côté, V. Binette, P. Tufféry, P. Derreumaux et N. Mousseau. Refining pep-fold for *do novo* peptide structure prediction with an all-atom extension of the coarse-grained potential opep. *In preparation.*, 2016.

[108] S. Côté, P. Derreumaux et N. Mousseau. Distinct morphologies for amyloid beta protein monomer : Aβ_{40}, aβ_{42} and aβ_{40}(d23n). *J. Chem. Theory Comput.*, 7:2584–2592, 2011.

[109] S. Côté, R. Laghaei, P. Derreumaux et N. Mousseau. Distinct dimerization for various alloforms of the amyloid-beta protein : Aβ_{1-40}, aβ_{1-42}, and aβ_{1-40}(d23n). *J. Phys. Chem. B*, 116(13):4043–4055, 2012.

[110] S. Côté, G. Wei et N. Mousseau. All-atom stability and oligomerization simulations of polyglutamine nanotubes with and without the 17-amino-acid n-terminal fragment of the huntingtin protein. *J. Phys. Chem. B*, 116:12168–12179, 2012.

[111] S. Côté, G. Wei et N. Mousseau. Atomistic mechanisms of huntingtin n-terminal fragment insertion on a phospholipid bilayer revealed by molecular dynamics simulations. *Proteins*, 82(7):1409–1427, 2014.

[112] Y. Crespo, F. Marinelli, F. Pietrucci et A. Laio. Metadynamics convergence law in a multidimensional system. *Phys. Rev. E*, 81:055701, 2010.

[113] S. L. Crick, M. Jayaraman, C. Frieden, R. Wetzel et R. V. Pappu. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *P. Natl. Acad. Sci. USA*, 103(45):16764–16769, 2006.

[114] S. L. Crick, K. M. Ruff, G. Kanchan, C. Frieden et Pappu R. V. Unmasking the roles of n- and c-terminal flanking sequences from exon 1 of huntingtin as modulators of polyglutamine aggregation. *Proc. Natl. Sci. USA*, 110(50):20075–20080, 2013.

[115] J. Danielsson, J. Jarvet, P. Damberg et A. GrÃd'slund. The alzheimer β-peptide shows temperature-dependent transitions between left-handed 3_1-helix, β-strand and random coil secondary structures. *FEBS J.*, 272:3938–3949, 2005.

[116] T. Darden, D. York et L. Pedersen. Particle mesh ewald : an n log(n) method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089–10092, 1993.

[117] G. Darnell, J. P. R. O. Orgel, R. Pahl et S. C. Meredith. Flanking polyproline sequences inhibit β-sheet structure in polyglutamine segments by inducing ppii-like helix structure. *J. Mol. Biol.*, 374:688–704, 2007.

[118] R. Das et D. Baker. Macromolecular modeling with rosetta. *Annu. Rev. Biochem.*, 77:363–382, 2008.

[119] X. Daura, R. Suter et van Gunsteren W. F. Validation of molecular simulation by comparison with experiment : rotational reorientation of tryptophan in water. *J. Chem. Phys.*, 110:3049–3055, 1999.

[120] S. W. Davies, M. Turmaine, B. A. Cozens, M. Difiglia, A. H. Sharp, C. A. Ross, E. Scherzinger, E. E. Wanker, L. Mangiarini et G. P. Bates. Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for hd mutation. *Cell*, 90:537–548, 1997.

[121] K. T. Debiec, A. M. Gronenborn et L. T. Chong. Evaluating the strength of salt bridges : a comparison of current biomolecular force fields. *J. Phys. Chem. B*, 118:6561–6569, 2014.

[122] P. Derreumaux. A diffusion process-controlled monte carlo method for finding the global energy minimum of a polypeptide chain. i. formulation and test on a hexadecapeptide. *J. Chem. Phys.*, 106:5260–5270, 1997.

[123] P. Derreumaux. Folding a 20 amino acid $\alpha\beta$ peptide with the diffusion process-controlled monte carlo method. *J. Chem. Phys.*, 107:1941–1947, 1997.

[124] P. Derreumaux. Finding the low-energy forms of avian pancreatic polypeptide with the diffusion-process-controlled monte carlo method. *J. Chem. Phys.*, 109(4):1567–1574, 1998.

[125] P. Derreumaux. Ab initio prediction of polypeptide structure from its sequence. *Comput. Phys. Commun.*, 139:121–122, 1999.

[126] P. Derreumaux. From polypeptide sequences to structures using monte carlo simulations and an optimized potential. *J. Chem. Phys.*, 111(5):2301–2310, 1999.

[127] P. Derreumaux. Generating ensemble averages for small proteins from extended conformations by monte carlo simulations. *Phys. Rev. Lett.*, 85(1):206–209, 2000.

[128] P. Derreumaux. Predicting helical hairpins from sequences by monte carlo simulations. *J. Comput. Chem.*, 21(7):582–589, 2000.

[129] P. Derreumaux. Insight into protein topology from monte carlo simulations. *J. Chem. Phys.*, 117:3499–3503, 2002.

[130] P. Derreumaux. Role of supersecondary structural elements in protein g folding. *J. Chem. Phys.*, 119(9): 4940–4944, 2003.

[131] P. Derreumaux et N. Mousseau. Coarse-grained protein molecular dynamics simulations. *J. Chem. Phys.*, 126(2):025101, 2007.

[132] C. J. Dickson, B. D. Madej, A. A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould et R. C. Walker. Lipid14 : The amber lipid force field. *J. Chem. Theory Comput.*, 10(2):865–879, 2014.

[133] M. DiFiglia, E. Sapp, K. Chase, C. Schwarz, A. Meloni, C. Young, E. Martin, J. P. Vonsattel, R. Carraway, S. A. Reeves, F. M. Boyce et N. Aronin. Huntingtin is a cytoplasmic protein associated with vesicles in human and rat brain neurones. *Neuron*, 14:1075–1081, 1995.

[134] M. DiFiglia, E. Sapp, K. O. Chase, S. W. Davies, G. P. Bates, J. P. Vonsattel et N. Aronin. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neuritis in brain. *Science*, 277:1990–1993, 1997.

[135] K. A. Dill et S. Bromberg. *Molecular driving forces : Statistical thermodynamics in biology, chemistry, physics and nanoscience.* Garland Science, Taylor and Francis Group, LLC, an inform business, 2e édition édition, 2010.

[136] K. A. Dill, S. B. Ozkan, T. R. Weikl, J. D. Chodera et V. A. Voelz. The protein folding problem : when will it be solved ? *Curr. Opin. Struc. Biol.*, 17:342–346, 2007.

[137] M. Dlugosz et J. Trylska. Secondary structures of native and pathogenic huntingtin n-terminal fragments. *J. Chem. Phys. B*, 115:11597–11608, 2011.

[138] X. Dong, W. Chen, N. Mousseau et P. Derreumaux. Energy landscapes of the monomer and dimer of the alzheimer's peptide aβ(1-28). *J. Chem. Phys.*, 128(12):125108, 2008.

[139] R. O. Dror, R. M. Dirks, J. P. Grossman, H. Xu et D. E. Shaw. Biomolecular simulation : a computational microscope for molecular biology. *Ann. Rev. Biophys.*, 41:429–452, 2012.

[140] R. O. Dror, T. J. Mildorf, D. Hilger, A. Manglik, D. W. Borhani, D. H. Arlow, A. Philippsen, N. Villanueva, Z. Yang, M. T. Lerch, W. L. Hubbell, B. K. Kobilka, R. K. Sunahara et D. E. Shaw. Structural basis for nucleotide exchange in heterotrimeric g proteins. *Science*, 348:1361–1365, 2015.

[141] Y. Duan, C. Wu, Chowdhury S., M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang et P. Kollman. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.*, 24:1999–2012, 2003.

[142] L. Dupuis et N. Mousseau. Understanding the ef-hand closing pathway using non-biased interatomic potentials. *J. Chem. Phys.*, 136:035101, 2012.

[143] M. P. Duyao, A. B. Auerbach, A. Ryan, F. Persichetti, G. T. Barnes, S. M. McNeil, J. P. Vonsattel, J. F. Gusella, A. L. Joyner et al. Inactivation of the mouse huntington's disease gene homolog hdh. *Science*, 269 (5222):407–410, 1995.

[144] D. J. Earl et M. W. Deem. Parallel tempering : theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.*, 7(23):3910–3916, 2005.

[145] P. Eastman, M. S. Friedrichs, J. D. Chodera, R. J. Radmer, C. M. Bruns, J. P. Ku, K. A. Beauchamp, T. J. Lane, L.-P. Wang, D. Shukla, T. Tye, M. Houston, T. Stich, C. Klein, M. R. Shirts et V. S. Pande. Openmm 4 : A reusable, extensive, hardware independent library for high performance molecular simulation. *J. Chem. Theory Comput.*, 9(1):461–469, 2013.

[146] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander et M. Scharf. The double cube lattice method : efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.*, 16:273–284, 1995.

[147] D. M. Engelman. Membranes are more mosaic than fluid. *Nature*, 438:578–580, 2005.

[148] W. P. Esler, E. R. Stimson, J. M. Jennings, H. V. Vinters, J. R. Ghilardi, J. P. Lee, P. W. Mantyh et J. E. Maggio. Alzheimer's disease amyloid propagation by a template-dependent dock-lock mechanism. *Biochemistry*, 39(21):6288–6295, 2000.

[149] L. Esposito, A. Paladino, C. Pedone et L. Vitagliano. Insights into structure, stability, and toxicity of monomeric and aggregated polyglutamine models form molecular dynamics simulations. *Biophys. J.*, 94:4031–4040, 2008.

[150] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee et L. Pedersen. A smooth particle mesh ewald potential. *J. Chem. Phys.*, 103:8577–8592, 1995.

[151] C. Eugene, N. Laghaei et N. Mousseau. Early oligomerization stages for the non-amyloid component of alpha-synuclein amyloid. *J. Chem. Phys.*, 141:135103, 2014.

[152] P. Faller. Copper and zinc binding to amyloid-β : coordination, dynamics, aggregation, reactivity and metal-ion transfer. *ChemBioChem*, 10:2837–2845, 2009.

[153] G. W. Feigenson. Phase behavior of lipid mixtures. *Nat. Chem. Biol.*, 2(11):560–563, 2006.

[154] A. Ferguson, Z. Liu et H. S. Chan. Desolvation barrier effects are a likely contributor to the remarkable diversity in the folding rates of small proteins. *J. Mol. Biol.*, 389:619–636, 2009.

[155] P. Ferrara, J. Apostolakis et A. Caflisch. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins*, 46:24–33, 2002.

[156] F. Fiumara, F. Luana, E. R. Kandel et W. A. Hendrickson. Essential role of coiled coils for aggregation and activity of q/n-rich prions and polyq proteins. *Cell*, 143:1121–1135, 2010.

[157] D. Flock, S. Colacino, G. Colombo et A. Di Nola. Misfolding of the amyloid β-protein : A molecular dynamics study. *Proteins*, 62:183–192, 2006.

[158] C. A. Flouda. Computational methods in protein structure prediction. *Biotechnol. Bioeng.*, 97(2):207–213, 2007.

[159] F. Forcellino et P. Derreumaux. Computer simulations aimed at structure prediction of super secondary motifs in proteins. *Proteins*, 45:159–166, 2001.

[160] N. K. Fox, S. E. Brenner et J.-M. Chandonia. Scope : Structural classification of proteins – extended, integrating scop and astral data and classification of new structures. *Nucl. Acids Res.*, 42:D304–D309, 2014.

[161] E.A. Fradinger, B.H. Monien, B. Urbanc, A. Lomakin, M. Tan, H. Li, S.M. Spring, M.M. Condron, L. Cruz, C.-W. Xie, G.B. Benedek et G. Bitan. C-terminal peptides coassemble into aβ42 oligomers and protect neurons against aβ42-induced neurotoxicity. *P. Natl. Acad. Sci. USA*, 105(37):14175–14180, 2008.

[162] P. L. Freddolino, F. Liu, M. Gruebele et K. Schulten. Ten-microsecond molecular dynamics simulation of a fast-folding ww domain. *Biophys. Lett.*, pages L75–L77, 2008.

[163] P. L. Freddolino, S. Park, B. Roux et K. Schulten. Force field bias in protein folding simulations. *Biophys. J.*, 96:3772–3780, 2009.

[164] D. Frishman et P. Argos. Knowledge-based protein secondary structure assignment. *Proteins*, 23:566–579, 1995.

[165] G. Fuertes, D. Gimenez, S. Esteban-Martin, A. J. Garica-Saez, O. Sanchez et J. Salgado. Roles of membrane lipids for the activity of pore forming peptides and proteins. Dans G. Anderluh et J. Lakey, éditeurs, *Proteins : membrane binding and pore formation.*, pages 31–55. Landes Bioscience and Springer Science+Business Media, 2010.

[166] H. Fukunishi, O. Watanabe et S. Takada. On the hamiltonian replica exchange method for efficient sampling of biomolecular systems : application to protein structure prediction. *J. Chem. Phys*, 116(20):9058–9067, 2002.

[167] A. E. Garcia et K. Y. Sanbonmatsu. α-helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci. USA*, 99:2782–2787, 2002.

[168] J. R. Gatchel et H. Y. Zoghbi. Diseases of unstable repeat expansion : mechanisms and common principles. *Nat. Rev. Genet.*, 6:743–755, 2005.

[169] M. M. Gessel, S. Bernstein, M. Kemper, D. B. Teplow et M. T. Bowers. Familial alzheimer's disease mutations differentially alter amyloid β-protein oligomerization. *ACS Chem. Neurosci.*, 3:909–918, 2012.

[170] M. L. Giuffrida, F. Caraci, B. Pignataro, S. Cataldo, P. De Bona, V. Bruno, G. Molinaro, G. Pappalardo, A. Messina, A. Palmigiano, D. Garozzo, F. Nicoletti, E. Rizzarelli et A. Copani. β-amyloid monomer are neuroprotective. *J. Neurosci.*, 29(34):10582–10587, 2009.

[171] L. Goerigk, C. A. Collyer et J. R. Reimers. Recommending hartee-fock theory with london-dispersion and basis-set-superposition corrections for the optimization or quantum refinement of protein structures. *J. Phys. Chem. B*, 118:14612–14626, 2014.

[172] T. J. Grabowski, H. S. Cho, J. P. G. Vonsattel, G. W. Rebeck et S. M. Greenberg. Novel amyloid precursor protein mutation in an iowa family with dementia and severe cerebral amyloid angiopathy. *Ann. Neurol.*, 49 (6):697–705, 2001.

[173] N. J. Greenfield. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.*, 1 (6):2876–2890, 2006.

[174] S. M. Gruner. Nonlamellar lipid phases. Dans P. L. Yeagle, éditeur, *The structure of biological membranes.*, pages 173–199. CRC Press LLC, 2004.

[175] X. Gu, E. R. Greiner, R. Mishra, R. Kodali, A. Osmand, S. Finkbeiner, J. S. Steffan, L. M. Thompson, R. Wetzel et X. W. Yang. Serines 13 and 16 are critical determinants of full-length human mutant huntingtin induced disease pathogenesis in hd mice. *Neuron*, 64:828–840, 2009.

[176] B. Guillot. A reappraisal of what we have learnt during three decades of computed simulations on water. *J. Mol. Liq.*, 101:219–260, 2002.

[177] F. Guyon et P. Tufféry. Assessing 3d scores for protein structure fragment mining. *Bioinformatics*, 2:67–77, 2010.

[178] F. Guyon et P. Tufféry. Fast protein fragment similarity scoring using a binet-cauchy kernel. *Bioinformatics*, 30:784–791, 2014.

[179] C. Haass et D. J. Selkoe. Soluble protein oligomers in neurodegeneration : lessons from the alzheimer's amyloid β-peptide. *Nat. Rev. Mol. Cell Biol.*, 8:101–112, 2007.

[180] S. L. Hands et A Wyttenbach. Neurotoxic protein oligomerization associated with polyglutamine diseases. *Acta Neuropathol.*, 120:419–437, 2010.

[181] U. H. E. Hansmann. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.*, 281:140–150, 1997.

[182] H. Hauser et G. Poupart. Lipid structure. Dans P. L. Yeagle, éditeur, *The structure of biological membranes*, pages 1–51. CRC Press LLC, 2004.

[183] M. Hauser, C. E. Mayer et J. Söding. kclust : fast and sensitive clustering of large protein sequence databases. *BMC Bioinformatics*, 14:248, 2013.

[184] B. Hess, H. Bekker, H. J. C. Berendsen et J. G. E. M. Fraaije. Lincs : a linear constraint solver for molecular simulations. *J. Comput. Chem.*, 18:1463–1472, 1997.

[185] B. Hess, C. Kutzner, D. van der Spoel et E. Lindahl. Gromacs 4 : algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4:435–447, 2008.

[186] Y. Hirakura, R. Azimov, R. Azimova et B. L. Kagan. Polyglutamine-induced ion channels : a possible mechanism for the neurotoxicity of huntington and other cag repeat diseases. *J. Neurosci. Res.*, 60:490–494, 2000.

[187] C. L. Hoop, H.-K. Lin, K. Kar, Z. Hou, M. A. Poirier, R. Wetzel et P. C. A. van der Wel. Polyglutamine amyloid core boundaries and flanking domain dynamics in huntingtin fragment fibrils determined by solid-state nuclear magnetic resonance. *Biochemistry*, 53:6653–6666, 2014.

[188] W. G. Hoover. Canonical dynamics : equilibrium phase-space distributions. *Phys. Rev. A*, 31(3):1695–1697, 1985.

[189] H. W. Horn, W. C. Swode, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura et T. Head-Gordon. Development of an improved four-site water model for biomolecular simulations : Tip4p-ew. *J. Chem. Phys.*, 120:9665–9678, 2004.

[190] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg et C. Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins*, 65:712–725, 2006.

[191] H. R. Horton, L. A. Moran, K. G. Scrimgeour, M. D. Perry et J. D. Rawn. *Principles of Biochemistry*. Pearson Prentice Hall, 4e édition édition, 2006.

[192] L. Hou, H. Shao, Y. Zhang, H. Li, N. K. Menon, E. B. Neuhaus, J. M. Brewer, I.-J. L. Byeon, D.G. Ray, M. P. Vitek, T. Iwashita, R. A. Makula, A. B. Przybyla et M. G. Zagorski. Solution nmr studies of the aβ(1-40) and aβ(1-42) peptides establish that the met35 oxidation state affects the mechanism of amyloid formation. *J. Am. Chem. Soc.*, 126(7):1992–2005, 2004.

[193] H. W. Huang. Elasticity of lipid bilayer interacting with amphiphilic helical peptides. *J. Phys. II France*, 5: 1427–1431, 1995.

[194] H. W. Huang. Molecular mechanism of peptide-induced pores in membranes. *Phys. Rev. Lett.*, 92:198304, 2004.

[195] A. Huet et P. Derreumaux. Impact of the mutation a21g (flemish variant) on alzheimer's β-amyloid dimers by molecular dynamics simulations. *Biophys. J.*, 91:3829–3840, 2006.

[196] International human genome sequencing consortium. Finishing the euchromatic sequence of the human genome. *Nature*, 431:931–945, 2004.

[197] A. Irbäck, S. Mitternacht et S. Mohanty. An effective all-atom potential for proteins. *PMC Biophys.*, 2:2, 2009.

[198] G. B. Irvine, O. M. El-Agnaf, G. M. Shankar et D. M. Walsh. Protein aggregation in the brain : the molecular basis for alzheimer's and parkinson's diseases. *Mol. Med.*, 14:451–464, 2008.

[199] J. Israelachvili.

[200] S. G. Itoh, H. Okumura et Y. Okamoto. Replica-exchange method in van der waals radius space : overcoming steric restrictions for biomolecules. *J. Chem. Phys.*, 132:134105, 2010.

[201] Martyna G. J., D. J. Tobias et M. L. Klein. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.*, 101(5):4177–4189, 1994.

[202] Martyna G. J., M. E. Tuckerman, D. J. Tobias et M. L. Klein. Explicit reversible integrators for extended systems dynamics. *Mol. Phys.*, 87(5):1117–1157, 1996.

[203] J. P. M. Jämbeck et A. P. Lyubartsev. Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. *J. Phys. Chem. B*, 116:3164–3179, 2012.

[204] J. P. M. Jämbeck et A. P. Lyubartsev. An extension and further validation of an all-atomistic force field for biological membrane. *J. Chem. Theory Comput.*, 8:2938–2948, 2012.

[205] J. P. M. Jämbeck et A. P. Lyubartsev. Another piece of the membrane puzzle : extending slip ids further. *J. Chem. Theory Comput.*, 9(1):774–784, 2013.

[206] J. P. M. Jämbeck et A. P. Lyubartsev. Exploring the free energy landscape of solutes embedded in lipid bilayers. *J. Phys. Chem. Lett.*, 4:1781–1787, 2013.

[207] H. Jang, F. Teran Arce, S. Ramachandran, R. Capone, R. Lai et R. Nussinov. Structural convergence among diverse, toxic beta-sheet ion channels. *J. Phys. Chem. B*, 114(29):9445–94451, 2010.

[208] H. Jang, J. Zheng et R. Nussinov. Models of beta-amyloid ion channels in the membrane suggest that channel formation in the bilayer is a dynamic process. *Biophys. J*, 93(6):1938–1949, 2007.

[209] M. Jayaraman, R. Kodali, B. Sahoo, A. K. Thakur, A. Mayasundari, R. Mishra, C. B. Peterson et R. Wetzel. Slow amyloid nucleation via α-helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments. *J. Mol. Biol.*, 415:881–899, 2012.

[210] M. Jayaraman, R. Mishra, R. Kodali, A. K. Thakur, L. M. I. Koharudin, Gronenborn A. M. et R. Wetzel. Kinetically competing huntingtin aggregation pathways control amyloid polymorphism and properties. *Biochemistry*, 51:2706–2716, 2012.

[211] Y. Jia, Z. Qian, Y. Zhang et G. Wei. Adsorption and orientation of human islet amyloid polypeptide (hiapp) monomer at anionic lipid bilayer : implications for membrane-mediated aggregation. *Int. J. Mol. Sci.*, 14(3): 6241–6258, 2013.

[212] F. Jiang, C.-Y. Zhou et Y.-D. Wu. Residue-specific force-field based on the protein coil library. rsff1 : modification of opls-aa\l. *J. Phys. Chem. B*, 118:6983–6998, 2014.

[213] W. L. Jorgensen. Transferable intermolecular potential functions for water, alcohols and ethers. application to liquid water. *J. Am. Chem. Soc.*, 103:335–340, 1981.

[214] W. L. Jorgensen. Revised tips for simulations of liquid water and aqueous solutions. *J. Chem. Phys.*, 77(7): 4156–4163, 1982.

[215] W. L. Jorgensen, J. Chandrasekhar et J. D. Madura. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935, 1983.

[216] W. L. Jorgensen, J. Chandrasekhar et J. D. Madura. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935, 1983.

[217] W. L. Jorgensen, D. S. Maxwell et J. Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118(45):11225–11236, 1996.

[218] W. L. Jorgensen et J. Tirado-Rives. The opls potential functions for proteins. energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, 110(6):1657–1666, 1988.

[219] B. L. Kagan, Y. Hirakura, R. Azimov et R. Azimova. The channel hypothesis of huntington's disease. *Brain Res. Bull.*, 56:281–284, 2001.

[220] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives et W. L. Jorgensen. Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B*, 105(28):6474–6487, 2001.

[221] Y. K. Kang et H. S. Park. Assessment of ccsd(t), mp2, dft-d, cbs-qb3 and g4(mp2) for conformational study of alanine and proline dipeptides. *Chem. Phys. Lett.*, 600:112–117, 2014.

[222] K. Kar, M. Jayaraman, B. Sahoo, R. Kodali et R. Wetzel. Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. *Nat. Struct. Mol. Biol.*, 18(2):328–337, 2011.

[223] P. Kar, S. M. Gopal, Y.-M. Cheng, A. Predeus et M. Feig. Primo : A transferable coarse-grained force field for proteins. *J. Chem. Theory Comput.*, 9(8):3769–3788, 2013.

[224] M. Karplus. Behind the folding funnel diagram. *Nat. Chem. Biol.*, 7:401–404, 2011.

[225] M. Karplus et D. H. Anderson. Valence-bond interpretation of electron-coupled nuclear spin interactions - application to methane. *J. Chem. Phys.*, 30:6–10, 1959.

[226] H. Kaya et H. S. Chan. Solvation effects and driving forces for protein thermodynamic and kinetic cooperativity : how adequate is native-centric topological modeling ? *J. Mol. Biol.*, 326:911–931, 2003.

[227] H. Kaya, Z. Liu et H. S. Chan. Chevron behavior and isostable enthalpic barriers in protein folding : successes and limitations of simple go-like modeling. *Biophys. J.*, 89:520–535, 2005.

[228] H. Kaya, Z. Uzunoglu et H. S. Chan. Spatial ranges of driving forces are a key determinant of protein folding cooperativity and rate diversity. *Phys. Rev. E*, 88:044701, 2013.

[229] R. Kayed et C. A. Lasagna-Reeves. Molecular mechanisms of amyloid oligomers toxicity. *J. Alzheimers Dis.*, 33:S67–S78, 2013.

[230] K. B. Kegel, E. Sapp, J. Alexander, A. Valencia, P. Reeves, X. Li, N. Masso, L. Sobin, N. Aronin et M. DiFiglia. Polyglutamine expansion in huntingtin alters its interaction with phospholipids. *J. Neurochem.*, 110: 1585–1597, 2009.

[231] K. B. Kegel, E. Sapp, J. Yoder, B. Cuiffo, L. Sobin, Y. J. Kim, Z.-H. Qin, M. R. Hayden, N. Aronin, D. L. Scott, G. Isenberg, W. H. Goldmann et M. DiFiglia. Huntingtin associates with acidic phospholipids at the plasma membrane. *J. Biol. Chem.*, 280:36464–36473, 2005.

[232] K. B. Kegel, V. Schewkunow, E. Sapp, N. Masso, E. E. Wanker, M. DiFiglia et W. H. Goldmann. Polyglutamine expansion in huntingtin increases its insertion into lipid bilayers. *Biochem. Biophys. Res. Co.*, 387: 472–475, 2009.

[233] N. W. Kelley, X. Huang, S. Tam, C. Spiess, J. Frydman et V. S. Pande. The predicted structure of the headpiece of the huntingtin protein and its implications on huntingtin aggregation. *J. Mol. Biol.*, 388(5): 919–927, 2009.

[234] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff et D. C. Phillips. A three-dimensional model of the myoglobin molecular obtained by x-ray analysis. *Nature*, 181:662–666, 1958.

[235] S. D. Khare, F. Ding, K. N. Gwanmesia et Dokholyan N. V. Molecular origin of polyglutamine aggregation in neurodegenerative diseases. *PLoS Comput. Biol.*, 1(3):e30, 2005.

[236] G. A. Khoury, A. Liwo, F. Khatib, H. Zhou, G. Chopra, J. Bacardit, L. O. Bortot, R. A. Faccioli, X. Deng, Y. He, P. Krupa, J. Li, M. A. Mozolewska, A. K. Sieradzan, J. Smadbeck, T. Wirecki, S. Cooper, J. Flatten, K. Xu, D. Baker, J. Cheng, A. C. B. Delbem, C. A. Floudas, C. Keasar, M. Levitt, Z. Popović, H. A. Scheraga, J. Skolnick, S. N. Crivelli et Foldit Players. Wefold : A coopetition for protein structure prediction. *Proteins*, 82:1850–1868, 2014.

[237] G. A. Khoury, J. Smadbeck, C. A. Kieslich et C. A. Flouda. Protein folding and de novo protein design for biotechnological applications. *Trends Biotechnol.*, 32:99–109, 2014.

[238] K. A. Kill et J. L. MacCallum. The protein-folding problem, 50 years on. *Science*, 338:1042–1046, 2012.

[239] M. W. Kim, Y. Chelliah, S. W. Kim, Z. Otwinowski et I. Bezprozvanny. Secondary structure of huntingtin amino-terminal region. *Structure*, 17(9):1205–1212, 2009.

[240] S. Kim, T. Takeda et D. K. Klimov. Mapping conformational ensembles of aβ oligomers in molecular dynamics simulations. *Biophys. J.*, 99(6):1949–1958, 2010.

[241] S. Kim, T. Takeda et D.K. Klimov. Globular state in the oligomers formed by aβ peptides. *J. Chem. Phys.*, 132(22):225101, 2010.

[242] Y. J. Kim, Y. Yi, E. Sapp, Y. Wang, B. Cuiffo, K. B. Kegel, Z.-H. Qin, N. Aronin et M. DiFiglia. Caspase 3-cleaved n-terminal fragments of wild-type and mutant huntingtin are present in normal and huntington's disease brains, associate with membranes, and undergo calpain-dependent proteolysis. *P. Natl. Acad. Sci. USA*, 98(22):12784–12789, 2001.

[243] M. D. Kirkitadze, M. M. Condron et D. B. Teplow. Identification and characterization of key kinetic intermediates in amyloid β-protein fibrillogenesis. *J. Mol. Biol.*, 312:1103–1119, 2001.

[244] J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, L. Vorobyov et A. D. MacKerell. Update of the charmm all-atom additive force field for lipids : validation on six lipid types. *J. Phys. Chem. B*, 114:7830–7843, 2010.

[245] F. A. C. Klein, A. Pastore, L. Masino, G. Zeder-Lutz, H. Nierengarten, M. Oulad-Abdelghani, D. Altschuh, J.-L. Mandel et Y. Trottier. Pathogenic and non-pathogenic polyglutamine tracts have similar structural properties : towards a length-dependent toxicity gradient. *J. Mol. Biol.*, 371:235–244, 2007.

[246] W. L. Klein. Synaptotoxic amyloid-beta oligomers : a molecular basis for the cause, diagnosis, and treatment of alzheimer's disease ? *J. Alzheimers Dis.*, 33:S49–S65, 2013.

[247] N. Koga, R. Tatsumi-Koga, G. Liu, R. Xiao, T. B. Acton, G. T. Montelione et D. Baker. Principles for designing ideal protein structures. *Nature*, 491:222–229, 2012.

[248] K. J. Kohlhoff, P. Robustelli, A. Cavalli, X. Salvatella et M. Vendruscolo. Fast and accurate predictions of protein nmr chemical shifts from interatomic distances. *J. Am. Chem. Soc.*, 131:13894–13895, 2009.

[249] P. Kollman, R. Dixon, W. Cornell, T. Fox, C. Chipot et A. Pohorille. The development/application of a 'minimalist' organic/biochemical molecular mechanic force field using a combination of *ab initio* calculations and experimental data. Dans van Gunsteren W. G., P. K. Weiner et A. J. Wilkinson, éditeurs, *Computer simulations of biomolecular systems*, pages 83–96. Kluwer Academic Publishers, 1997.

[250] E. V. Koonin, Y. I. Wolf et G. P. Karev. The structure of the protein universe and genome evolution. *Nature*, 420:218–223, 2002.

[251] V. Kräntler, W. F. van Gunsteren et P. H. Hünenberger. A fast shake algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.*, 22:501–508, 2001.

[252] G. G. Krivov, M. V. Shapovalov et R. L. Dunbrack Jr. Improved prediction of protein side-chain conformations with scwrl4. *Proteins*, 77:778–795, 2009.

[253] B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard et D. Baker. Design of a novel globular protein fold with atomic-level accuracy. *Science*, 302:1364–1368, 2003.

[254] S. Kumar-Singh, J. Theuns, B. Van Broeck, D. Pirici, K. Vennekens, E. Corsmit, M. Cruts, B. Dermaut, R. Wang et C. Van Broeckhoven. Mean age-of-onset of familial alzheimer disease caused by presenilin mutations correlates with both increased both increased aβ_{42} and decreased aβ_{40}. *Hum. Mutat.*, 27:686–695, 2006.

[255] R. Laghaei et N. Mousseau. Spontaneous formation of polyglutamine nanotubes with molecular dynamics simulations. *J. Chem. Phys.*, 132:165102, 2010.

[256] R. Laghaei, N. Mousseau et G. Wei. The effect of disulfide bond on the monomeric structure of human amylin studied by combined hamiltonian and temperature replica exchange molecular dynamics simulations. *J. Phys. Chem. B.*, 114:7071–7077, 2010.

[257] R. Laghaei, N. Mousseau et G. Wei. Structure and thermodynamics of amylin dimer studied by hamiltonian-temperature replica exchange molecular dynamics simulations. *J. Phys. Chem. B*, 115:3146–3154, 2011.

[258] A. Laio et F. L. Gervasio. Metadynamics : a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.*, 71:126601, 2008.

[259] A. Laio et M. Parrinello. Escaping free-energy minima. *P. Natl. Acad. Sci. USA*, 99:12562–12566, 2002.

[260] V. V. Lakhani, F. Ding et N. V. Dokholyan. Polyglutamine induced misfolding of huntingtin exon1 is modulated by the flanking sequences. *PLoS Comput. Biol.*, 6(4):e1000772, 2010.

[261] A. R. Lam, D. B. Teplow, H. E. Stanley et B. Urbanc. Effects of the arctic ($e^{22} \rightarrow g$) mutation on amyloid β-protein folding : Discrete molecular dynamics study. *J. Am. Chem. Soc.*, 130:17413–17422, 2008.

[262] T. J. Lane, D. Shukla, K. A. Beauchamp et V. S. Pande. To millisecond and beyond : challenges in the simulation of protein folding. *Curr. Opin. Struc. Biol.*, 23:58–65, 2013.

[263] H. A. Lashuel et S. M. Butterfield. Amyloidogenic protein-membrane interactions : mechanistic insight from model systems. *Angew. Chem. Int. Ed.*, 49:5628–5654, 2010.

[264] H. A. Lashuel et P. T. Jr. Lansbury. Are amyloid diseases caused by protein aggregates that mimic bacterial pore-forming toxins ? *Q. Rev. Biophys.*, 39(2):167–201, 2006.

[265] N. D. Lazo, M. A. Grant, M. C. Condron, A. C. Rigby et D. B. Teplow. On the nucleation of amyloid β-protein monomer folding. *Protein Sci.*, 14:1581–1596, 2005.

[266] C. C. Lee, R. H. Walters et R. M. Murphy. Reconsidering the mechanism of polyglutamine peptide aggregation. *Biochemistry*, 46:12810–12820, 2007.

[267] M. Levitt, M. Gerstein, E. Huang, S. Subbiah et J. Tsai. Protein folding : The endgame. *Annu. Rev. Biochem.*, 66:549–579, 1997.

[268] M. Levitt et S. Lifson. Refinement of protein conformations using a macromolecular energy minimization procedure. *J. Mol. Biol.*, 46(46):269–279, 1969.

[269] M. Levitt et R. Sharon. Accurate simulation of protein dynamics in solution. *Proc. Natl. Acad. Sci. USA*, 85:7557–7561, 1988.

[270] M. Levitt et A. Warshel. Computer simulation of protein folding. *Nature*, 253(5494):94–96, 1975.

[271] R. Lewis et R. N. McElhaney. The mesomorphic phase behavior of lipid bilayers. Dans P. L. Yeagle, éditeur, *The structure of biological membranes*, pages 53–120. CRC Press LLC, 2004.

[272] D.-W. Li et Brüschweiler. Nmr-based protein potentials. *Angew. Chem. Int. Ed.*, 49:6778–6780, 2010.

[273] D.-W. Li et Brüschweiler. Iterative optimization of molecular mechanics force fields from nmr data of full-length proteins. *J. Chem. Theory Comput.*, 7:1773–1782, 2011.

[274] C. Liang, P. Derreumaux, N. Mousseau et G. Wei. The beta-strand-beta-strand conformation is marginally populated in the $beta_2$-microglobulin (20-41) peptide in solution as revealed by replica exchange molecular dynamics simulations. *Biophysical J.*, 95:510–517, 2008.

[275] S. W. Liebman et S. C. Meredith. Sticky n17 speeds huntingtin pile-up. *Nat. Chem. Biol.*, 6:7–8, 2010.

[276] S. Lifson et A. Warshel. Consistent force field for calculations of conformations, vibrational spectra, and enthalpies of cycloalkane and n-alkane molecules. *J. Chem. Phys.*, 49(11):5116–5129, 1968.

[277] K. H. Lim, H. H. Collver, Y. T. H. Le, P. Nagchowdhuri et J. M. Kenney. Characterizations of distinct amyloidogenic conformations of the aβ(1-40) and (1-42) peptides. *Biochem. Biophys. Res. Co.*, 353:443–449, 2007.

[278] E. Lindahl, B. Hess et D. van der Spoel. Gromacs 3.0 : a package for molecular simulation and trajectory analysis. *J. Mol. Mod.*, 7:306–317, 2001.

[279] K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror et D. E. Shaw. Systematic validation of protein force fields against experimental data. *PLoS ONE*, 7(2):e32131, 2012.

[280] K. Lindorff-Larsen, S. Piana, R. O. Dror et D. E. Shaw. How fast-folding proteins fold. *Science*, 334: 517–520, 2011.

[281] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror et D. E. Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins*, 78(8):1950–1958, 2010.

[282] D. Lingwood et K. Simons. Lipids rafts as a membrane-organizing principle. *Science*, 327:46–50, 2010.

[283] R. A. Lippert, C. Predescu, D. J. Ierardi, K. M. Mackenzie, M. P. Eastwood, R. O. Dror et D. E. Shaw. Accurate and efficient integration for molecular dynamics simulations at constant temperature and pressure. *J. Chem. Phys.*, 139:164106, 2013.

[284] R. Liu, C. McAllister, Y. Lyubchenko et M. Sierks. Residues 17-20 and 30-35 of beta-amyloid play critical roles in aggregation. *J. Neurosci. Res.*, 75:162–171, 2004.

[285] Z. Liu et H. S. Chan. Desolvation is a likely origin of robust enthalpic barriers to protein folding. *J. Mol. Biol.*, 349:872–889, 2005.

[286] Z. Liu et H. S. Chan. Solvation and desolvation effects in protein folding : native flexibility, kinetic cooperativity and enthalpic barriers under isostability conditions. *Phys. Biol.*, 2:S75–S85, 2005.

[287] A. Liwo, C. Czaplewski, S. Oldziej, A. V. Rojas, R. Kazmierkiewicz, M. Makowski, R. K. Murarka et H. A. Scheraga. Simulation of protein structure and dynamics with the coarse-grained unres force field. Dans G. A. Voth, éditeur, *Coarse-graining of condensed phase and biomolecular systems.*, pages 1391–1411. CRC Press, 2008.

[288] C. Lockhart, S. Kim, R. Kumar et D. K. Klimov. Does amino acid sequence determine the properties of aβ dimer ? *J. Chem. Phys.*, 135(3):035103, 2011.

[289] S. C. Lovell, I. W. Davis, W. Bryan Arendall III, P. I. W. de Bakker, J. Michael Word, M. G. Prisant, J. S. Richardson et D. C. Richardson. Structure validation by cα geometry : ϕ, ψ and cβ deviation. *Proteins*, 50: 437–450, 2003.

[290] Y. Lu, P. Derreumaux, Z. Guo, N. Mousseau et Guanghong. Wei. Thermodynamics and dynamics of amyloid peptide oligomerisation are sequence-dependent. *Proteins*, 75:954–963, 2009.

[291] Y. Lu, G. Wei et P. Derreumaux. Effects of g33a and g33i mutations on the structures of monomer and dimer of the amyloid-β fragment 29-42 by replica exchange molecular dynamics simulations. *J. Phys. Chem. B*, 115:1282–1288, 2011.

[292] Y. Lu, G. Wei et P. Derreumaux. Structural, thermodynamical, and dynamical properties of oligomers formed by the amyloid nnqq peptide : Insights from coarse-grained simulations. *J. Chem. Phys.*, 137:025101, 2012.

[293] A. L. Lublin et S. Gandy. Amyloid-β oligomers : possible roles as key neurotoxins in alzheimer's disease. *Mt. Sinai J. Med.*, 77:43–49, 2010.

[294] T. Luhrs, C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Dobeli, D. Schubert et R. Riek. 3d structure of alzheimer's amyloid-β(1-42) fibrils. *P. Natl. Acad. Sci. USA*, 102(48):17342–17347, 2005.

[295] A. P. Lyubartsev et A Laaksonen. Mdynamix – a scalable portable parallel md simulation package for arbitrary molecular mixtures. *Comput. Phys. Commun.*, 128:565–589, 2000.

[296] Fändrich M. Oligomeric intermediates in amyloid formation : structure determination and mechanisms of toxicity. *J. Mol. Biol.*, 421:427–440, 2012.

[297] B. Ma et R. Nussinov. Polymorphic triple beta-sheet structures contribute to amide hydrogen/deuterium (h/d) exchange protection in the alzheimer amyloid beta 42 peptide. *J. Biol. Chem.*, 286(39):34244–34253, 2011.

[298] M. E. MacDonald. Huntingtin : alive and well and working in middle management. *Sci. STKE*, 207:pe48, 2003.

[299] J. A. D. MacKerell, D. Bashford, R. L. Dunbrack, J. D. Evanseck, M. J. Fischer et al. All-atom empirical potential for molecular modelling and dynamics studies of proteins. *J. Phys. Chem. B*, 102(18):3586–3616, 1998.

[300] J. A. D. MacKerell, M. Feig et C. L. Brooks III. Extending the treatment of backbone energetics in protein force fields : limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.*, 25(11):1400–1415, 2004.

[301] M. W. Mahoney et W. L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.*, 112(20):8910–8922, 2000.

[302] T. Maiuri, T. Woloshansky, J. Xia et R. Truant. The huntingtin n17 domain is a multifunctional crm1 and rand-dependent nuclear and cilia export signal. *Hum. Mol. Genet.*, 22(7):1383–1394, 2013.

[303] S. K. Maji, R. R. O. Loo, M. Inayathullah, S. M. Spring, S. S. Vollers, Condron M. M. et al. Amino acid position-specific contributions to amyloid β-protein assembly and toxicity. *J. Biol. Chem.*, 284:23580–23591, 2009.

[304] L. Mangiarini, K. Sathasivam, M. Seller, B. Cozens, A. Harper, C. Hetherington, M. Lawton, Y. Trottier, H. Lehrach, S. W. Davies et G. P. Bates. Exon 1 of the hd gene with an expanded cag repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell*, 87:493–506, 1996.

[305] A. J. Marchut et C. K. Hall. Side-chain interactions determine amyloid formation by model polyglutamine peptides in molecular dynamics simulations. *Biophys. J.*, 90:4574–4584, 2006.

[306] A. J. Marchut et C. K. Hall. Effects of chain length on the aggregation of model polyglutamine peptides : molecular dynamics simulations. *Proteins*, 66:96–109, 2007.

[307] S. J. Marrink et D. P. Tieleman. Perspective on the martini model. *Chem. Soc. Rev.*, 42:6801–6822, 2013.

[308] G. J. Martyna, M. L. Klein et M. Tuckerman. Nosé-hoover chains : The canonical ensemble via continuous dynamics. *J. Chem. Phys.*, 97(3):2635–2643, 1992.

[309] J. Maupetit, P. Derreumaux et P. Tuffery. Pep-fold : an online resource for de novo peptide structure prediction. *Nucleic Acids Res.*, 37:W498–503, 2009.

[310] J. Maupetit, P. Derreumaux et P. Tuffery. A fast method for large-scale de novo peptide and miniprotein structure prediction. *J. Comput. Chem.*, 31(4):726–738, 2010.

[311] J. Maupetit, P. Tuffery et P. Derreumaux. A coarse-grained protein force field for folding and structure prediction. *Proteins*, 69:394–408, 2007.

[312] J. A. McCammon, B. R. Gelin et M. Karplus. Dynamics of folded proteins. *Nature*, 267:585–590, 1977.

[313] W. C. McMurray. Phospholipids in sub cellular organelles and membranes. Dans G. B. Ansell, J. N. Hawthorne et R. M. C. Dawson, éditeurs, *Form and function of phospholipids*, pages 205–251. Elsevier, 1973.

[314] A. Melquiond, X. Dong, N. Mousseau et P. Derreumaux. Role of the region 23-28 in aβ fibril formation : Insights from simulations of the monomers and dimers of alzheimer's peptides aβ40 and aβ42. *Curr. Alzheimer Res.*, 5:244–250, 2008.

[315] A. Melquiond, N. Mousseau et P. Derreumaux. Structures of soluble amyloid oligomers from computer simulations. *Proteins*, 65(1):180–91, 2006.

[316] D. Meral et B. Urbanc. Discrete molecular dynamics study of oligomer formation by n-terminally truncated amyloid beta-protein. *J. Mol. Biol.*, 425(12):2260–2275, 2012.

[317] M. Michalek, C. Aisenbrey et B. Bechinger. Investigation of membrane penetration depth and interactions of the amino-terminal domain of huntingtin : refined analysis by tryptophan fluorescence measurement. *Eur. Biophys. J.*, 2014.

[318] M. Michalek, E. S. Salnikov et B. Bechinger. Structure and topology of the huntingtin 1-17 membrane anchor by a combined solution and solid-state nmr approach. *Biophys. J.*, 105:699–710, 2013.

[319] M. Michalek, E. S. Salnikov, S. Werten et B. Bechinger. Membrane interactions of the amphipathic amino terminus of huntingtin. *Biochem. J.*, 52:847–858, 2013.

[320] M. S. Miettinen, V. Knecht, L. Monticelli et Z. Ignatova. Assessing polyglutamine conformation in the nucleating event by molecular dynamics simulations. *J. Phys. Chem. B*, 116:10259–10265, 2012.

[321] Y. Miller, B. Ma, C.-J. Tsai et R. Nussinov. Hollow core of alzheimer's aβ_{42} amyloid observed by cryoem is relevant at physiological ph. *P. Natl. Acad. Sci. USA*, 107(32):14128–14133, 2010.

[322] A. Mitsutake, Y. Mori et Y. Okamoto. Enhanced sampling algorithms. Dans L. Monticelli et E. Salonen, éditeurs, *Biomolecular simulations.*, pages 153–195. Humana Press, 2013.

[323] J. Mittal et R. B. Best. Tackling force-field bias in protein folding simulations : folding of billion hp35 and pin ww domains in explicit water. *Biophys. Lett.*, 99:L26–L28, 2010.

[324] S. Mitternacht, I. Staneva, T. Hard et A. Irback. Comparing the folding free-energy landscapes of aβ42 variants with different aggregation properties. *Proteins*, 78:2600–2608, 2010.

[325] S. Mitternacht, I. Staneva, T. Hard et A. Irback. Monte carlo study of the formation and conformational properties of dimers of aβ_{42} variants. *J. Mol. Biol.*, 410(2):357–367, 2011.

[326] S. Miyamoto et P. A. Kollman. Settle : an analytical version of the shake and rattle algorithm for rigid water models. *J. Comput. Chem.*, 13(8):952–962, 1992.

[327] H. Monoi, S. Futaki, S. Kugimiya, H. Minakata et K. Yoshihara. Poly-l-glutamine forms cation channels : relevance to the pathogenesis of the polyglutamine diseases. *Biophys. J.*, 78(6):2892–2899, 2000.

[328] C. Morgan, M. Colombres, M. T. Nunez et N. C. Inestrosa. Structure and function of amyloid in alzheimer's disease. *Prog. Neurobiol.*, 74:323–349, 2004.

[329] T. Mori, F. Ogushi et Y. Sugita. Analysis of lipid surface area in protein-membrane systems combining voronoi tessellation and monte carlo integration methods. *J. Comput. Chem.*, 55:379–400, 2012.

[330] J. Moult, K. Fidelis, A. Kryshtafovych, T. Schwede et A. Tramontano. Critical assessment of methods of protein structure prediction (casp) – round x. *Proteins*, 82:1–6, 2014.

[331] J. Moult, J. T. Pedersen, R. Judson et K. Fidelis. A large-scale experiment to assess protein structure prediction methods. *Proteins*, 23:ii–iv, 1995.

[332] N. Mousseau et P. Derreumaux. Exploring the early steps of amyloid peptide aggregation by computers. *Acc. Chem. Res.*, 38:885–891, 2005.

[333] N. Mousseau et P. Derreumaux. Exploring energy landscape of protein folding and aggregation. *Front. Biosci.*, 13:4495–4516, 2008.

[334] N. Mousseau, P. Derreumaux, G. T. Barkema et R. Malek. Sampling activated mechanisms in proteins with the activation-relaxation technique. *J. Mol. Graph. Model.*, 19:78–86, 2001.

[335] A. Nagarajan, S. Jawahery et S. Matysiak. The effects of flanking sequences in the interaction of polyglutamine peptides with a membrane bilayer. *J. Phys. Chem. B*, 118(24):6368–6379, 2014.

[336] M. Nakano, H. Watanabe, E. B. Starikov, S. M. Rothstein et S. Tanaka. Mutations effects on structural stability of polyglutamine peptides by molecular dynamics simulation. *Interdiscip. Sci. Comput. Life Sci.*, 1: 21–29, 2009.

[337] J. Nasica-Labouze, M. Meli, P. Derreumaux, G. Colombo et N. Mousseau. A multiscale approach to characterize the early aggregation steps of the amyloid-forming peptide gnnqqny from the yeast prion sup-35. *PLoS Comput. Biol.*, 7(5):e1002051, 2011.

[338] J. Nasica-Labouze et N. Mousseau. Kinetics of amyloid aggregation : A study of the gnnqqny prion sequence. *PLoS Comput. Biol.*, 8:e1002782, 2012.

[339] J. Nasica-Labouze et N. Mousseau. Kinetics of amyloid growth. Dans P. Derreumaux, éditeur, *Alzheimer's disease : Insights into low molecular weight and cytotoxic aggregates from in vitro and computer experiments.*, pages 209–237. Imperial College Press, 2013.

[340] J. Nasica-Labouze, P. H. Nguyen, F. Sterpone, O. Berthoumieu, N.-V. Buchete, S. Côté, A. De Simone, A. J. Doig, P. Faller, A. Garcia, A. Laio, M. S. Li, S. Melchionna, N. Mousseau, Y. Mu, A. Paravastu, S. Pasquali, D. J. Rosenman, B. Strodel, B. Tarus, J. H. Viles, T. Zhang, C. Wang et P. Derreumaux. Amyloid β protein and alzheimer's disease : When computer simulations complement experimental studies. *Chem. Rev.*, 115 (9):3518–3563, 2015.

[341] P. S. Nerenberg et T. Head-Gordon. Optimizing protein–solvent force fields to reproduce intrinsic conformational preferences of model peptides. *J. Chem. Theory Comput.*, 7:1220–1230, 2011.

[342] P. S. Nerenberg, B. Jo, C. So, A. Tripathy et T. Head-Gordon. Optimizing solute-water van der waals interactions to reproduce solvation free energies. *J. Phys. Chem. B*, 116:4524–4534, 2012.

[343] M. E. J. Newman et G. T. Barkema. Monte carlo methods in statistical physics. *Clarendon Press*, page 496 pages, Dec 1 1998.

[344] H. Nguyen, J. Maier, H. Huang, V. Perrone et C. Simmerling. Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. *J. Am. Chem. Soc.*, 136:13959–13962, 2014.

[345] P. H. Nguyen, M. S. Li et P. Derreumaux. Amyloid oligomer structure characterization from simulations : a general method. *J. Chem. Phys.*, 140:094105, 2014.

[346] P. H. Nguyen, Y. Okamoto et P. Derreumaux. Simulated tempering with fast on-the-fly weight determination. *J. Chem. Phys.*, 138(6):061102, 2013.

[347] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81(1):511–519, 1984.

[348] S. Nosé et M. L. Klein. Constant pressure molecular dynamics for molecular system. *Mol. Phys.*, 50(5): 1055–1076, 1983.

[349] K. Nozaki, O. Osamu, T. Hiroki et T. Shoji. Amino acids sequences flanking polyglutamine stretches influence their potential for aggregate formation. *Neuroreport*, 12(15):3357–3364, 2001.

[350] L. G. Nucifora, K. A. Burke, X. Feng, N. Arbez, S. Zhu, J. Miller, G. Yang, T. Ratovitski, M. Delannoy, P. J. Muchowski, S. Finkbeiner, J. Legleiter, C. A. Ross et M. A. Poirier. Identification of novel potentially toxic oligomers formed in vitro from mammalian-derived expanded huntingtin exon-1 protein. *J. Biol. Chem.*, 287(19):16017–16028, 2012.

[351] E. P. O'Brien, Y. Okamoto, J. E. Straub, B. R. Brooks et D. Thirumalai. Thermodynamic perspective on the dock-lock growth mechanism of amyloid fibrils. *J. Phys. Chem. B*, 113(43):14421–14430, 2009.

[352] H. Ogawa, M. Nakano, H. Watanabe, E. B. Starikov, S. M. Rothstein et S. Tanaka. Molecular dynamics simulation study on the structural stabilities of polyglutamine peptides. *Comput. Biol. Chem.*, 32:102–110, 2008.

[353] C. J. Oldfield et A. K. Dunker. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.*, 83:553–584, 2014.

[354] A. Olofsson, M. Lindhagen-Persson, A. E. Sauer-Eriksson et A. Ohman. Amide solvent protection analysis demonstrates that amyloid-β(1-40) and amyloid-β(1-42) form different fibrillar structures under identical conditions. *Biochem. J.*, 404:63–70, 2007.

[355] K. Ono, M. M. Condron et D. B. Teplow. Structure-neurotoxicity relationships of amyloid β-protein oligomers. *P. Natl. Acad. Sci. USA*, 106(35):14745–14750, 2009.

[356] K. Ono, M. M. Condron et D. B. Teplow. Effects of the english (h6r) and tottori (d7n) familial alzheimer disease mutations on amyloid β-protein assembly and toxicity. *J. Biol. Chem.*, 285(30):23186–23197, 2010.

[357] C. Oostenbrink, A. Villa, A. E. Mark et W. F. Van Gunsteren. A biomolecular force field field based on the free enthalpy of hydration and solvation : the gromos force-field parameter sets 53a5 and 53a6. *J. Comput. Chem.*, 25(13):1656–1676, 2004.

[358] H. T. Orr et H. Y. Zoghbi. Trinucleotide repeat disorders. *Annu. Rev. Neurosci.*, 30:575–621, 2007.

[359] Ulmschneider J. P., J. C. Smith, M. B. Ulmschneider, A. S. Ulrich et E. Strandberg. Reorientation and dimerization of the membrane-bound antimicrobial peptide pgla from microsecond all-atom md simulations. *Biophys. J.*, 103:472–482, 2012.

[360] C. Paissoni, D. Spiliotopoulos, G. Musco et A. Spitaleri. Gmxpbsa 2.0 : A gromacs tool to perform mm/pbsa and computational alanine scanning. *Comput. Phys. Comm.*, 185:2920–2929, 2014.

[361] M. Paloncýová, G. Fabre, R. H. DeVane, P. Trouillas, K. Berka et M. Otyepka. Benchmarking of force fields for molecule-membrane interactions. *J. Chem. Theory Comput.*, 10:4143–4151, 2014.

[362] J. Pan, J. Han, C. H. Borchers et L. Konermann. Conformer-specific hydrogen exchange analysis of aβ(1-42) oligomers by top-down electron capture dissociation mass spectrometry. *Anal. Chem.*, 83(13):5386–5393, 2011.

[363] J. Pan, J. Han, C. H. Borchers et L. Konermann. Structure and dynamics of small soluble aβ(1-40) oligomers studied by top-down hydrogen exchange mass spectrometry. *Biochemistry*, 51(17):3694–3703, 2012.

[364] J. Pan, T. T. Mills, S. Tristram-Nagle et J. F. Nagle. Cholesterol perturbs lipid bilayers nonuniversally. *Phys. Rev. Lett.*, 100:198103, 2008.

[365] V. S. Pande, I. Baker, J. Chapman, S. P. Elmer, S. Khaliq, S. M. Larson, Y. M. Rhee, M. R. Shirts, C. D. Snow, E. J. Sorin et B. Zagrovic. Atomistic protein folding simulations on the sub millisecond time scale using worldwide distributed computing. *Biopolymers*, 68(1):91–109, 2003.

[366] A. K. Paravastu, R. D. Leapman, W.-M. Yau et R. Tycko. Molecular structural basis for polymorphism in alzheimer's β-amyloid fibrils. *Proc. Natl. Acad. Sci. USA*, 105:18349–18354, 2008.

[367] A. K. Paravastu, A. T. Petkova et R. Tycko. Polymorphic fibril formation by residues 10-40 of the alzheimer's β-amyloid peptide. *Biophys. J.*, 90(12):4618–4629, 2006.

[368] M. Parrinello et A. Rahman. Polymorphic transitions in single crystals : a new molecular dynamics method. *J. Appl. Phys.*, 52:7182, 1981.

[369] M. Pasi, R. Lavery et N. Ceres. Palace : A coarse-grained protein model for studying mechanical properties. *J. Chem. Theory Comput.*, 9(1):785–793, 2013.

[370] S. Pasquali et P. Derreumaux. Hire-rna : A high resolution coarse-grained energy model for rna. *J. Phys. Chem. B*, 114:11957–11966, 2010.

[371] A. Patriksson et D. van der Spoel. A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.*, 10:2073–2077, 2008.

[372] T. Perevozchikova, C. B. Stanley, H. P. McWilliams-Koeppen, E. L. Rowe et V. Berthelier. Investigating the structural impact of the glutamine repeat in huntingtin assembly. *Biophys. J.*, 107:411–421, 2014.

[373] M. F. Perutz. Glutamine repeats as polar zippers : their role in inherited neurodegenerative disease. *Mol. Med.*, 1(7):718–721, 1995.

[374] M. F. Perutz, J. T. Finch, J. Berriman et A. Lesk. Amyloid fibers are water-filled nanotubes. *P. Natl. Acad. Sci. USA*, 99(8):5591–5595, 2002.

[375] A. T. Petkova, Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio et R. Tycko. A structural model for alzheimer's β-amyloid fibrils based on experimental constraints from solid state nmr. *P. Natl. Acad. Sci. USA*, 99(26):16742–16747, 2002.

[376] A. T. Petkova, R. D. Leapman, Z. Guo, W.-M. Yau, M. P. Mattson et R. Tycko. Self-propagating, molecular-level polymorphism in alzheimer's β-amyloid fibrils. *Science*, 307:262–265, 2005.

[377] J. C. Philipps, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé et K. Schulten. Scalable molecular dynamics with namd. *J. Comp. Chem.*, 26(16):1781–1802, 2005.

[378] R. Phillips, T. Ursell, P. Wiggins et P. Sens. Emerging roles for lipids in shaping membrane-protein function. *Nature*, 459:379–385, 2009.

[379] S. Piana, A. G. Donchev, P. Robustelli et D. E. Shaw. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B*, 119:5113–5123, 2015.

[380] S. Piana, J. L. Klepeis et D. E. Shaw. Assessing the accuracy of physical models used in protein-folding simulations : quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struc. Biol.*, 24: 98–105, 2014.

[381] S. Piana et A. Laio. A bias-exchange approach to protein folding. *J. Phys. Chem. B*, 111:4553–4559, 2007.

[382] S. Piana, K. Lindorff-Larsen et D. E. Shaw. How robust are protein folding simulations with respect to force field parametrization ? *Biophys. J.*, 100:L47–L49, 2011.

[383] S. Piana, K. Lindorff-Larsen et D. E. Shaw. Protein folding kinetics and thermodynamics from atomistic simulation. *P. Natl. Acad. Sci. USA*, 109(44):17845–17850, 2012.

[384] T. J. Piggot, Á. Piñeiro et S. Khalid. Molecular dynamics simulations of phosphatidylcholine membranes : a comparative force field study. *J. Chem. Theory Comput.*, 8:4593–4609, 2012.

[385] A. A. Polyansky, A. O. Chugunov, A. A. Vassilevski, E. V. Grishin et R. G. Efremov. Recent advances in computational modelling of alpha-helical membrane-active peptides. *Curr. Protein Peptide Sci.*, 13:644–657, 2012.

[386] M. K. Prakash, A. Barducci et M. Parrinello. Replica temperatures for uniform exchange and efficient roundtrip times in explicit solvent parallel tempering simulations. *J. Chem. Theory Comput.*, 7(7):2025–2027, 2011.

[387] S. Pronk, S. Pali, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess et E. Lindahl. Gromacs 4.5 : a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, 2013.

[388] D. F. Raffa et A. Rauk. Molecular dynamics study of the beta amyloid peptide of alzheimer's disease and its divalent copper complexes. *J. Phys. Chem. B*, 111(14):3789–3799, 2007.

[389] G. N. Ramachandran, C. Ramakrishnan et V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.*, 7:95–99, 1963.

[390] A. Ramamoorthy et M. H. Lim. Structural characterization and inhibition of toxic amyloid-beta oligomeric intermediates. *Biophys. J.*, 105(2):287–288, 2013.

[391] R. P. Rand et V. A. Parsegian. Hydration forces between phospholipid bilayers. *Biochim. Biophys. Acta*, 988:351–376, 1989.

[392] T. Ratovistski, M. Nakamura, J. D'Ambola, E. Chighladze, Y. Liang, W. Wang, R. Graham, M. R. Hayden, D. R. Borchelt, R. R. Hirschhorn et C. A. Ross. N-terminal proteolysis of full-length mutant huntingtin in an inducible pc12 cell model of huntington's disease. *Cell Cycle*, 6(23):2970–2981, 2007.

[393] W. Rawicz, B. A. Smith, T. J. McIntosh, S. A. Simon et E. Evans. Elasticity strength and water permeability of bilayers that contain raft microdomain-forming lipids. *Biophys. J.*, 94:4725–4736, 2008.

[394] G. Reddy, J. E. Straub et D. Thirumalai. Influence of preformed asp23-lys28 salt bridge one the conformational fluctuations of monomers and dimers of aβ peptides with implications for rates of fibril formation. *J. Phys. Chem. B*, 113(4):1162–1172, 2009.

[395] J. J. Ritch, A. Valencia, J. Alexander, E. Sapp, L. Gatune, G. R. Sangrey, S. Sinha, C. M. Scherber, S. Zeitlin, G. Sadri-Vakili, D. Irimia, M. DiFiglia et K. B. Kegel. Multiple phenotypes in huntington disease mouse neural stem cells. *Mol. Cell. Neurosci.*, 50(1):70–81, 2012.

[396] A. L. Robertson, M. A. Bate, A. M. Buckle et S. P. Bottomley. The rate of polyq-mediated aggregation is dramatically affected by the number and location of surrounding domains. *J. Mol. Biol.*, 413:879–887, 2011.

[397] M. J. Robertson, J. Tirado-Rives et W. L. Jorgensen. Improved peptide and protein torsional energetics with the opls-aa force field. *J. Chem. Theory Comput.*, 11:3499–3509, 2015.

[398] E. Rockabrand, N. Slepko, A. Pantalone, V. N. Nukala, A. Kazantsev, J. L. Marsh, P. G. Sullivan, J. S. Steffan, S. L. Sensi et L. M. Thompson. The first 17 amino acids of huntingtin modulate its sub-cellular localization, aggregation and effects on calcium homeostasis. *Hum. Mol. Genet.*, 16(1):61–77, 2007.

[399] C. A. Ross et M. A. Poirier. What is the role of protein aggregation in neurodegeneration ? *Nat. Rev. Mol. Cell Bio.*, 6:891–898, 2005.

[400] C. A. Ross, M. A. Poirier, E. E. Wanker et M. Amzel. Polyglutamine fibrillogenesis : the pathway unfolds. *P. Natl. Acad. Sci. USA*, 100(1):1–3, 2003.

[401] G. Rossetti, P. Cossio, A. Laio et P. Carloni. Conformations of the huntingtin n-term in aqueous solution from atomistic simulations. *FEBS Lett.*, 585:3086–3089, 2011.

[402] R. Roychaudhuri, M. Yang, A. Deshpande, G. M. Cole, S. Frautschy, A. Lomakin et al. C-terminal turn stability determines assembly differences between abeta40 and abeta42. *J. Mol. Biol.*, 425(2):292–308, 2013.

[403] K. M. Ruff, S. J. Khan et Pappu R. V. A coarse-grained model for polyglutamine aggregation modulated by amphipathic flanking sequences. *Biophys. J.*, 107(5):1226–1235, 2014.

[404] J.-P. Ryckaert, G. Ciccotti et H. J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints : Molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23:327–341, 1977.

[405] B. Sahoo, D. Singer, R. Kodali, T. Zuchner et R. Wetzel. Aggregation behavior of chemically synthesized, full-length huntingtin exon 1. *Biochemistry*, 53:3897–3907, 2014.

[406] M. Sakono et T. Zako. Amyloid oligomers : formation and toxicity of aβ oligomers. *FEBS J.*, 277:1348–1358, 2010.

[407] A. Saladin, J. Rey, P. Thévenet, M. Zacharias, G. Moroy et P. Tuffery. Pep-sitefinder : a tool for the blind identification of peptide binding sites on protein surfaces. *Nucleic Acids Res.*, 42:W221–W226, 2014.

[408] E. Salnikov, P. Bertani, J. Raap et B. Bechinger. Analysis of the amide ^{15}n chemical shift tensor of the c$_\alpha$ tetra substituted constituent of membrane-active peptaibols, the α-aminoisobutyric acid residue compared to those of di- and tri-substituted proteinogenic amino acid residues. *J. Biomol. NMR*, 45:373–387, 2009.

[409] H. M. Saunders et S. P. Bottomley. Multi-domain misfolding : understanding the aggregation pathway of polyglutamine proteins. *J. Mol. Biol.*, 22(8):447–451, 2009.

[410] T. Schlick. *Molecular modeling and simulation : An interdisciplinary guide.* Springer, 2e édition édition, 2010.

[411] T. Schlick, R. Collepardo-Guevara, L. A. Halvorsen, S. Jung et X. Xiao. Biomolecular modelling and simulation : a field coming of age. *Q. Rev. Biophys.*, 44(2):191–228, 2011.

[412] J. M. Schmidt, M. Blumel, F. Lohr et H. Ruterjans. Self-consistent ^{3}j coupling analysis for the joint calibration of karplus coefficients and evaluation of torsion angles. *J. Biomol. NMR*, 14:1–12, 1999.

[413] T. H. Schmidt et C. Kandt. Lambada and inflategro2 : efficient membrane alignment and insertion of membrane proteins for molecular dynamics simulations. *J. Chem. Inf. Model*, 52:2657–2669, 2012.

[414] R. Schneider, M. C. Schumacher, H. Mueller, D. Nand, V. Klaukien, H. Heise, D. Riedel, G. Wolf, E. Behrmann, S. Raunser, R. Seidel, M. Engelhard et M. Baldus. Structural characterization of polyglutamine fibrils by solid-state nmr spectroscopy. *J. Mol. Biol.*, 412:121–136, 2011.

[415] S. Schwarzinger, G. J. A. Kroon, T. R. Foss, J. Chung, P. E. Wright et H. J. Dyson. Sequence-dependent correction of random coil nmr chemical shifts. *J. Am. Chem. Soc.*, 123(13):2970–2978, 2001.

[416] K. L. Sciarretta, D. J. Gordon, A. T. Petkova, R. Tycko et S. C. Meredith. Aβ-lactam(d23/k28) models a conformation highly favorable for nucleation of amyloid. *Biochemistry*, 44(16):6003–6014, 2005.

[417] J. D. Selkoe. Folding proteins in fatal ways. *Nature*, 426:900–904, 2003.

[418] J. D. Selkoe. Soluble oligomers of the amyloid β-protein impair synaptic plasticity and behavior. *Behav. Brain Res.*, 192(1):106–113, 2008.

[419] N. G. Sgourakis, M. Merced-Serrano, C. Boutsidis, P. Drineas, Z. Du, C. Wang et A. E. Garcia. Atomic-level characterization of the ensemble of the aβ_{42} monomer in water using unbiased molecular dynamics simulations and spectral algorithms. *J. Mol. Biol.*, page In press, 2010.

[420] N. G. Sgourakis, Y. Yan, S. McCallum, C. Wang et A. E. Garcia. The alzheimer's peptides aβ40 and 42 adopt distinct conformations in water : A combined md/nmr study. *J. Mol. Biol.*, 368(5):1448–1457, 2007.

[421] Y. Shafrir, S. R. Durell, A. Anishkin et H. R. Guy. Beta-barrel models of soluble amyloid beta oligomers and annular protofibrils. *Proteins*, 78(16):3458–3472, 2010.

[422] G. M. Shankar, S. Li, T. H. Mehta, A. Garcia-Munoz, N. E. Shepardson, I. Smith, F. M. Brett, M. A. Farrell, M. J. Rowan, C. A. Lemere, C. M. Regan, D. M. Walsh, B. L. Sabatini et D. J. Selkoe. Amyloid β-protein dimers isolated directly from alzheimer brains impair synaptic plasticity and memory. *Nat. Med.*, 14(8): 837–842, 2008.

[423] D. Sharma, L. M. Shinchuk, H. Inouye, R. Wetzel et D. A. Kirschner. Polyglutamine homopolymers having 8 to 45 residues form slablike β-crystallite assemblies. *Proteins*, 61:398–411, 2005.

[424] A. H. Sharp, S. J. Loev, G. Schilling, S. H. Li, X. J. Li, J. Bao, M. V. Wagster, J. A. Kotzuk, J. P. Steiner, A. Lo, J. Hedreen, S. Sisodia, S. H. Snyder, T. M. Dawson, D. K. Ryugo et C. A. Ross. Widespread expression of huntington's disease gene (it15) product. *Neuron*, 14(5):1065–1074, 1995.

[425] D. E. Shaw, R. O. Dror, J. K. Salmon, J. P. Grossmna, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, E. Chow, M. P. Eastwood, D. J. Ierardi, J. S. Klepeis, J. L. Kuskin, R. H. Larson, K. Lindorff-Larsen, P. Maragakis, M. A. Moraes, S. Piana, Y. Shan et B. Towles. Millisecond-scale molecular dynamics simulations on anton. Dans *Proceedings of the international conference for high performance computing, networking, storage and analysis.*, page 39. ACM, 2009.

[426] D. E. Shaw, J. P. Grossman, J. A. Bank, B. Batson, J. A. Butts, J. C. Chao, M. M. Deneroff, R. O. Dror, A. Even, C. H. Fenton, A. Forte, J. Gagliardo, G. Gill, B. Greskamp, C. R. Ho, J. D. Ierardi, J. S. Iserovich, L. Kuskin, R. H. Larson, T. Layman, L.-S. Lee, A. K. Lerer, C. Li, D. Killebrew, K. M. Mackenzie, S. Y.-H. Mok, M. A. Moraes, R. Mueller, L. J. Nociolo, J. L. Peticolas, T. Quan, D. Ramot, J. K. Salmon, D. P. Scarpazza, U. B. Schafer, N. Siddique, C. W. Snyder, J. Spengler, P. T. P. Tang, M. Theobald, H. Toma, B. Towles, B. Vitale, S. C. Wang et C. Young. Anton 2 : Raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. Dans T. Damkroger et J. Dongarra, éditeurs, *Proceedings of the international conference for high performance computing, networking, storage and analysis.*, pages 41–53. IEEE, 2014.

[427] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan et W. Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330:341–346, 2010.

[428] J.-E. Shea et B. Urbanc. Insights into aβ aggregation : a molecular dynamics perspective. *Curr. Top. Med. Chem.*, 12:2596–2610, 2012.

[429] Y. Shen et A. Bax. Sparta+ : a modest improvement in empirical nmr chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR*, 48:13–22, 2010.

[430] Y. Shen, J. Maupetit, P. Derreumaux et P. Tufféry. Improved pep-fold approach for peptide and miniprotein structure prediction. *J. Chem. Theory Comput.*, 10(10):4745–4758, 2014.

[431] D. P. Siegel. Lipid membrane fusion. Dans P. L. Yeagle, éditeur, *The structure of biological membranes.*, pages 255–308. CRC Press LLC, 2004.

[432] P. Sikorski et E. Atkins. New model for crystalline polyglutamine assemblies and their connection with amyloid fibrils. *Biomacromolecules*, 6:425–432, 2005.

[433] S. J. Singer et G. L. Nicolson. The fluid mosaic model of the structure of cell membranes. *Science*, 175: 720–731, 1972.

[434] V. N. Sivanandam, M. Jayaraman, C. L. Hoop, R. Kodali, R. Wetzel et P. C. A. van der Wel. The aggregation-enhancing huntingtin n-terminus is helical in amyloid fibrils. *J. Am. Chem. Soc.*, 133(12):4558–4566, 2011.

[435] E. Socher, H. Sticht et A. H. C. Horn. The conformational stability of nonfibrillar amyloid-β peptide oligomers critically depends on the c-terminal peptide length. *ACS Chem. Neurosci.*, 5(3):161–167, 2014.

[436] E. J. Sorin et V. S. Pande. Empirical force-field assessment : the interplay between backbone torsions and non covalent term scaling. *J. Comput. Chem.*, 26:682–690, 2005.

[437] E. J. Sorin et V. S. Pande. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys. J.*, 88:2472–2493, 2005.

[438] C. Soto. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat. Rev. Neurosci.*, 4: 49–60, 2003.

[439] Y. G. Spill, S. Pasquali et P. Derreumaux. Impact of thermostats on folding and aggregation properties of peptides using the optimized potential for efficient structure prediction coarse-grained model. *J. Chem. Theory Comput.*, 7:1502–1510, 2011.

[440] D. V. D. Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark et H. J. C. Berendsen. Gromacs : fast, flexible, and free. *J. Comput. Chem.*, 26:1701–1718, 2005.

[441] J.-F. St-Pierre et N. Mousseau. Large loop conformation sampling using the activation relaxation technique, art-nouveau method. *Proteins*, 80:1883–1894, 2012.

[442] J.-F. St-Pierre, N. Mousseau et P. Derreumaux. The complex folding pathways of protein a suggest a multiple-funnelled energy landscape. *J. Chem. Phys.*, 128:045101, 2008.

[443] J. S. Steffan, N. Agrawal, J. Pallos, E. Rockabrand, L. C. Trotman, N. Slepko, K. Illes, T. Lukacsovich, Y.-Z. Zhu, E. Cattaneo, P. P. Pandolfi, L. M. Thompson et J. L. Marsh. Sumo modification of huntingtin and huntington's disease pathology. *Science*, 304:100–104, 2004.

[444] F. Sterone, S. Melchionna, P. Tuffery, S. Pasquali, N. Mousseau, T. Cragnolini, Y. Chebaro, J.-F. St-Pierre, M. Kalimeri, A. Barducci, Y. Laurin, A. Tek, M Baaden, P. H. Nguyen et P. Derreumaux. The opep protein model : from single molecules, amyloid formation, crowing and hydrodynamics to dna/rna systems. *Chem. Soc. Rev.*, 43:4871–4893, 2014.

[445] F. Sterpone, P. H. Nguyen, M. Kalimeri et P. Derreumaux. Importance of the ion-pair interactions in the opep coarse-grained force-field : parametrization and validation. *J. Chem. Theory Comput.*, 9:4574–4584, 2013.

[446] M. Stork, A. Giese, H. A. Kretzschmar et P. Tavan. Molecular dynamics simulations indicate a possible role of parallel β-helices in seeded aggregation of poly-gln. *Biophys. J.*, 88:2442–2451, 2005.

[447] V. A. Streltsov, J. N. Varghese, C. L. Masters et S. D. Nuttall. Crystal structure of the amyloid-β p3 fragment provides a model for oligomer formation in alzheimer's disease. *J. Neurosci.*, 31:1419–1426, 2011.

[448] Y. Sugita, A. Kitao et Y. Okamoto. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.*, 113(15):6042–6051, 2000.

[449] Y. Sugita et Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314:141–151, 1999.

[450] L. Sutto, S. Marsili et F. L. Gervasio. New advances in metadynamics. *WIREs Comput. Mol. Sci.*, 2:771–779, 2012.

[451] Y. Suzuki, J. R. Brender, M. T. Soper, J. Krishnamoorthy, Y. Zhou, B. T. Ruotolo et al. Resolution of oligomeric species during the aggregation of aβ1-40 using ^{19}f nmr. *Biochemistry*, 52(11):1903–1912, 2013.

[452] W. C. Swope, H. C. Andersen, Berens P. H. et K. R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules : Application to small water clusters. *J. Chem. Phys.*, 76:637–649, 1982.

[453] T. Takeda et D. K. Klimov. Interpeptide interactions induce helix to strand structural transitions in aβ peptides. *Proteins*, 77:1–13, 2009.

[454] T. Takeda et D. K. Klimov. Probing energetics of aβ fibril elongation by molecular dynamics simulations. *Biophys. J.*, 96(11):4428–4437, 2009.

[455] T. Takeda et D. K. Klimov. Probing the effect of amino-terminal truncation for aβ_{1-40} peptides. *J. Phys. Chem. B*, 113(19):6692–6702, 2009.

[456] T. Takeda et D. K. Klimov. Computational backbone mutagenesis of aβ peptides : probing the role of backbone hydrogen bonds in aggregation. *J. Phys. Chem. B*, 114:4755–4762, 2010.

[457] S. Tam, C. Spiess, W. Auyeung, L. Joachimiak, B. Chen, M. A. Poirier et J. Frydman. The chaperoning tric blocks a huntingtin sequence element that promotes the conformational switch to aggregation. *Nat. Struct. Mol. Biol.*, 16(12):1279–1286, 2009.

[458] W. M. Tay, D. Huang, T. L. Rosenberry et A. K. Paravastu. The alzheimer's amyloid-beta(1-42) peptide forms off-pathway oligomers and fibrils that are distinguished structurally by intermolecular organization. *J. Mol. Biol.*, 425(14):2494–2508, 2013.

[459] A. K. Thakur, M. Jayaraman, R. Mishra, M. Thakur, V. M. Chellgren, I.-J. L. Byeon, D. H. Anjum, R. Kodali, T. P. Creamer, J. F. Conway, A. M. Gronenborn et R. Wetzel. Polyglutamine disruption of the huntingtin exon 1 n terminus triggers a complex aggregation mechanism. *Nat. Struct. Mol. Biol.*, 16(4):380–389, 2009.

[460] A. K. Thakur et R. Wetzel. Mutational analysis of the structural organization of polyglutamine aggregates. *P. Natl. Acad. Sci. USA*, 99(26):17014–17019, 2002.

[461] P. Thévenet, J. Rey, G. Moroy et P. Tuffery. De novo peptide structure prediction : an overview. *Methods Mol. Biol.*, 1268:1–13, 2015.

[462] P. Thévenet, Y. Shen, J. Maupetit, F. Guyon, P. Derreumaux et P. Tuffery. Pep-fold : an updated *de novo* structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res.*, 40: W288–W293, 2012.

[463] L. M. Thompson, C. T. Aiken, L. S. Kaltenbach, N. Agrawal, K. Illes, A. Khoshnan, M. Martinez-Vincente, M. Arrasate, J. G. O'Rourke, H. Khashwji, T. Lukacsovich, Y.-Z. Zhu, A. L. Lau, A. Massey, M. R. Hayden, S. O. Zeitlin, S. Finkbeiner, K. N. Green, F. M. LaFerla, G. Bates, L. Huang, P. H. Patterson, D. C. Lo, A. M. Cuervo, J. L. Marsh et J. S. Steffan. Ikk phosphorylates huntingtin and targets it for degradation by the proteasome and lysosome. *J. Cell Biol.*, 187:1083–1099, 2009.

[464] J. T. Titantah et M. Karttunen. Water dynamics : Relation between hydrogen bond bifurcations, molecular jumps, local density and hydrophobicity. *Sci. Rep.*, 3:2991, 2013.

[465] P. Tiwary et M. Parrinello. A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B*, 119:736–742, 2015.

[466] S. Trebst, M. Troyer et U. H. E. Hansmann. Optimized parallel tempering simulations of proteins. *J. Chem. Phys.*, 124:174903, 2006.

[467] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni et G. Bussi. Plumed 2 : new feathers for an old bird. *Comput. Phys. Commun.*, 185:604–613, 2014.

[468] L. Triguero, R. Singh et R. Prabhakar. Comparative molecular dynamics studies of wild-type and oxidized forms of full-length alzheimer amyloid β-peptides aβ(1-40) and aβ(1-42). *J. Phys. Chem. B*, 112(23): 7123–7131, 2008.

[469] R Tycko. Molecular structure of amyloid fibrils : insights from solid-state nmr. *Q. Rev. Biophys.*, 39(1): 1–55, 2006.

[470] R Tycko. . *Annu. Rev. Phys. Chem.*, 62(Solid-state NMR studies of amyloid fibril structure.):279–299, 2011.

[471] R. Tycko, K. L. Sciarretta, J. P. R. O. Orgel et S. C. Meredith. Evidence for novel β-sheet structures in iowa mutant β-amyloid fibrils. *Biochemistry*, 48(26):6072–6084, 2009.

[472] S. Uppamoochikkal, S. Tristram-Nagle et J. F. Nagle. Orientation of tie-lines in the phase diagram of dopc/dppc/cholesterol model biomembranes. *Langmuir*, 26:17363–17368, 2010.

[473] B. Urbanc, M. Betnel, L. Cruz, G. Bitan et D.B. Teplow. Elucidation of amyloid β-protein oligomerization mechanisms : Discrete molecular dynamics study. *J. Am. Chem. Soc.*, 132:4266–4280, 2010.

[474] B. Urbanc, J. M. Borreguero, L. Cruz et H. E. Stanley. Ab initio discrete molecular dynamics approach to protein folding and aggregation. *Methods Enzymol.*, 412:314–338, 2006.

[475] B. Urbanc, L. Cruz, G. Ding, D. Sammond, S. Khare, S. V. Buldyrev et al. Molecular dynamics simulation of amyloid beta dimer formation. *Biophys. J.*, 87(4):2310–2321, 2004.

[476] B. Urbanc, L. Cruz, S. Yun, S. V. Buldyrev, G. Bitan, D. B. Teplow et H. E. Stanley. In silico study of amyloid β-protein folding and oligomerization. *P. Natl. Acad. Sci. USA*, 101(50):17345–17350, 2004.

[477] R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright et M. M. Babu. Classification of intrinsically disordered regions and proteins. *Chem. Rev.*, 114: 6589–6631, 2014.

[478] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark et H. J. C. Berendsen. Gromacs : fast, flexible and free. *J. Comput. Chem.*, 26:1701–1719, 2005.

[479] C. Vega et J. L. F. Abascal. Simulating water with rigid non-polarizable models : a general perspective. *Phys. Chem. Chem. Phys.*, 13:19663–19688, 2011.

[480] M. Vendruscolo et C. M. Dobson. Protein dynamics : Moore's law in molecular biology. *Curr. Biol.*, 21(2): R68–R70, 2011.

[481] L. Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159(1):98–103, 1967.

[482] M. H. Viet, P. H. Nguyen, S. T. Ngo, M. S. Li et P. Derreumaux. Effect of the tottori familial disease mutation (d7n) on the monomers and dimers of aβ40 and aβ42. *ACS Chem. Neurosci.*, 4(11):1446–1457, 2013.

[483] A. Vitalis, N. Lyle et R. V. Pappu. Thermodynamics of β-sheet formation in polyglutamine. *Biophys. J.*, 97: 303–311, 2009.

[484] A. Vitalis et R. V. Pappu. Assessing the contribution of heterogeneous distributions of oligomers to aggregation mechanisms of polyglutamine peptides. *Biophys. Chem.*, 159:14–23, 2011.

[485] A. Vitalis, X. Wang et R. V. Pappu. Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization. *J. Mol. Biol.*, 384(1): 279–297, 2008.

[486] J. Vreede, M. G. Wolf, S. W. de Leeuw et P. G. Bolhuis. Reordering hydrogen bonds using hamiltonian replica exchange enhances sampling of conformational changes in biomolecular systems. *J. Chem. Phys. B*, 113:6484–6494, 2009.

[487] G. W. Vuister et A. Bax. Quantitative j correlation – a new approach for measuring homonuclear 3-bond j(h(n)h(alpha)) coupling-constants in n-15-enriched proteins. *J. Am. Chem. Soc.*, 115:7772–7777, 1993.

[488] D. M. Walsh et D. B. Teplow. Alzheimer's disease and the amyloid beta-protein. *Prog. Mol. Biol. Transl.*, 107:101–124, 2012.

[489] D.M. Walsh et D. J. Selkoe. Aβ oligomers - a decade of discovery. *J. Neurochem.*, 101:1172–1184, 2007.

[490] J. Wang, P. Cieplak et P. A. Kollman. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules ? *J. Comput. Chem.*, 21(12): 1049–1074, 2000.

[491] J.-S. Wang et R. H. Swendsen. Replica monte-carlo simulation of spin-glasses. *Phys. Rev. Lett.*, 57:2607, 1986.

[492] L. Wang, R. A. Friesner et B. J. Berne. Replica exchange with solute scaling : a more efficient version of replica exchange with solute tempering (rest2). *J. Phys. Chem. B*, 115(30):9431–9438, 2011.

[493] F. H. C. Watson, J. D. et Crick. Molecular structure of nucleic acids. *Nature*, 171(4356):737–738, 1953.

[494] G. Wei, P. Derreumaux et N. Mousseau. Sampling the complex energy landscape of a simple β-hairpin. *J. Chem. Phys.*, 119(13):6403–6406, 2003.

[495] G. Wei, N. Mousseau et P. Derreumaux. Exploring the energy landscape of proteins : A characterization of the activation-relaxation technique. *J. Chem. Phys.*, 117:11379–11387, 2002.

[496] G. Wei, N. Mousseau et P. Derreumaux. Complex folding pathways in a simple β-hairpin. *Proteins*, 56: 464–474, 2004.

[497] G. Wei, N. Mousseau et P. Derreumaux. Characterizing the early steps of protein aggregation : structures and pathways. *Prion*, 1(1):3–8, 2007.

[498] G. Wei, W. Song, P. Derreumaux et N. Mousseau. Self-assembly of amyloid-forming peptides by molecular dynamics simulations. *Front. Biosci.*, 13:5681–5692, 2008.

[499] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Jr. Profeta et P. Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, 106: 765–784, 1984.

[500] S. J. Weiner, P. A. Kollman, D. T. Nguyen et D. A. Case. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.*, 7(2):230–252, 1986.

[501] R. Wetzel. Physical chemistry of polyglutamine : Intriguing tales of a monotonous sequence. *J. Mol. Biol.*, 421:466–490, 2012.

[502] D. A. White. The phospholipid composition of mammalian tissues. Dans G. B. Ansell, J. N. Hawthorne et R. M. C. Dawson, éditeurs, *Form and function of phospholipids*, pages 441–482. Elsevier, 1973.

[503] M. P. Williamson, T. F. Havel et K. Wüthrich. Solution conformation of proteinase inhibitor iia from bull seminal plasma by ^{1}h nuclear magnetic resonance and distance geometry. *J. Mol. Biol.*, 182:295–315, 1985.

[504] T. E. Williamson, A. Vitalis, S. L. Crick et R. V. Pappu. Modulation of polyglutamine conformations and dimer formation by the n-terminus of huntingtin. *J. Mol. Biol.*, 396(5):1295–1309, 2010.

[505] W. C. Wimley. Energetics of peptide and protein binding to lipid membranes. *Proteins : membrane binding and pore formation*, 13:14–23, 2010.

[506] A. Wimo et M. Prince. World alzheimer report 2010 – the global economic impact of dementia. *Alzheimer's Disease International*, page 9, 2010.

[507] P. G. Wolynes, J. N. Onuchic et D. Thirumalai. Navigating the folding routes. *Science*, 267:1619–1620, 1995.

[508] P. E. Wright et H. J. Dyson. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell. Biol.*, 16:18–29, 2015.

[509] J. Xu et Y. Zhang. How significant is a protein structure similarity with tm-score=0.5 ? *Bioinformatics*, 26: 889–895, 2010.

[510] Y. Xu, J. Shen, X. Luo, W. Zhu, K. Chen, J. Ma et H. Jiang. Conformational transition of amyloid β-peptide. *P. Natl. Acad. Sci. USA*, 102(15):5403–5407, 2005.

[511] Y. Yan et C. Wang. Aβ42 is more rigid than aβ40 at the c terminus : Implications for aβ aggregation and toxicity. *J. Mol. Biol.*, 364:853–862, 2006.

[512] M. Yang et D. B. Teplow. Amyloid β-protein monomer folding : Free-energy surfaces reveal alloform-specific differences. *J. Mol. Biol.*, 384:450–464, 2008.

[513] A. Yano, A. Okamoto, K. Nomura, S. Higai et N. Kurita. Difference in dimer conformation between amyloid-β(1-42) and (1-43) proteins : replica-exchange molecular dynamics simulations in water. *Chem. Phys. Lett.*, 595-596:242–249, 2014.

[514] P. L. Yeagle. The roles of cholesterol in the biology of cells. Dans P. L. Yeagle, éditeur, *The structure of biological membranes.*, pages 243–254. CRC Press LLC, 2004.

[515] L. Yu, R. Edalji, J. E. Harlan, T. F. Holzman, A. P. Lopez, B. Labkovsky et al. Structural characterization of a soluble amyloid beta-peptide oligomer. *Biochemistry*, 48:1870–1877, 2009.

[516] X. Yu et J. Zheng. Polymorphic structures of alzheimer's β-amyloid globulomers. *PLoS ONE*, 6(6):e20575, 2011.

[517] S. Yun, B. Urbanc, L. Cruz, G. Bitan, D. B. Teplow et H. E. Stanley. Role of electrostatic interactions in amyloid beta-protein (aβ) oligomer formation : a discrete molecular dynamics study. *Biophys. J.*, 92(11): 4064–4077, 2007.

[518] S. Yun, S. Yun et H. R. Guy. Analysis of the stabilities of hexameric amyloid-β(1-42) models using discrete molecular dynamics simulations. *J. Mol. Graph. Model.*, 29(5):657–662, 2011.

[519] Lv. Z., R. Roychaudhuri, M. M. Condron, D. B. Teplow et Y. L. Lyubchenko. Mechanism of amyloid beta-protein dimerization determined using single-molecule afm force spectroscopy. *Sci. Rep.*, 3:2880, 2013.

[520] D. Zanuy, K. Gunasekaran, A. M. Lesk et R. Nussinov. Computational study of the fibril organization of polyglutamine repeats reveals a common motif identified in β-helices. *J. Mol. Biol.*, 358:330–345, 2006.

[521] A. Zemel, A. Ben-Shaul et S. May. Membrane perturbation induced by interracially adsorbed peptides. *Biophys. J.*, 86:3607–3619, 2004.

[522] A. Zemel, A. Ben-Shaul et S. May. Perturbation of a lipid membrane by amphipathic peptides and its role in pore formation. *Eur. Biophys. J.*, 34:230–242, 2005.

[523] A. Zhang, W. Qi, T. A. Good et E. J. Fernandez. Structural differences between aβ(1-40) intermediate oligomers and fibrils elucidated by proteolytic fragmentation and hydrogen/deuterium exchange. *Biophys. J.*, 96(3):1091–1104, 2009.

[524] C. Zhang et J. Ma. Folding helical proteins in explicit solvent using dihedral-biased tempering. *Proc. Natl. Acad. Sci. USA*, 109(21):8139–8144, 2012.

[525] M. Zhang, J. Zhao et J. Zheng. Molecular understanding of a potential functional link between anti microbial and amyloid peptides. *Soft Matter*, 10:7425–7451, 2014.

[526] Q. C. Zhang, T.-L. Yeh, A. Leyva, L. G. Frank, J. Miller, Y. E. Kim, R. Langen, S. Finkbeiner, M. L. Amzel, C. A. Ross et M. A. Poirier. A compact β model of huntingtin toxicity. *J. Biol. Chem.*, 286(10):8188–8196, 2011.

[527] R. Zhang, X. Hu, H. Khant, S. J. Ludtke, W. Chiu, M. F. Schmid, C. Frieden et J.-M. Lee. Interprotofilament interactions between alzheimer's aβ_{1-42} peptides in amyloid fibrils revealed by cryoem. *P. Natl. Acad. Sci. USA*, 106(12):4653–4658, 2009.

[528] S. Zhang, K. Iwata, M. J. Lachenmann, J. W. Peng, S. Li, E. R. Stimson, Y.-a. Lu, A. M. Felix, J. E. Maggio et J. P. Lee. The alzheimer's peptide aβ adopts a collapsed coil structure in water. *J. Struct. Biol.*, 130: 130–141, 2000.

[529] T. Zhang, J. Zhang, P. Derreumaux et Y. Mu. Molecular mechanism of the inhibition of egcg on the alzheimer aβ(1-42) dimer. *J. Phys. Chem. B*, 117(15):3993–4002, 2013.

[530] Y. Zhang. Progress and challenges in protein structure prediction. *Curr. Opin. Struc. Biol.*, 18:342–348, 2008.

[531] Y. Zhang, Y. Luo, Y. Deng, Y. Mu et G. Wei. Lipid interaction and membrane perturbation of human islet amyloid polypeptide monomer and dimer by molecular dynamics simulations. *PLoS ONE*, 7:e38191, 2012.

[532] Y. Zhang, D. L. Rempel, J. Zhang, A. K. Sharma, L. M. Mirica et M. L. Gross. Pulsed hydrogen-deuterium exchange mass spectrometry probes conformational changes in amyloid beta (aβ) peptide aggregation. *Proc. Natl. Acad. Sci. USA*, 110(36):14604–14609, 2013.

[533] Y. Zhang et J. Skolnick. Scoring function for automated assessment of protein structure template quality. *Proteins*, 57:702–710, 2004.

[534] J. Zheng, H. Jang, B. Ma et R. Nussinov. Annular structures as intermediates in fibril formation of alzheimer aβ17-42. *J. Phys. Chem. B*, 112:6856–6865, 2008.

[535] Z. Zheng et M. I. Diamond. Huntington disease and the huntingtin protein. *Prog. Mol. Biol.*, 107:189–214, 2012.

[536] Z. Zheng, A. Li, B. B. Holmes, J. C. Marasa et M. I. Diamond. An n-terminal nuclear export signal regulates trafficking and aggregation of huntingtin (htt) protein exon 1. *J. Biol. Chem.*, 288(9):6063–6071, 2013.

[537] Z. Zheng, A. Li, B. B. Holmes, J. C. Marasa et M. I. Diamond. An n-terminal nuclear export signal regulates trafficking and aggregation of huntingtin (htt) protein exon 1. *J. Biol. Chem.*, 288(9):6063–6071, 2013.

[538] X. Zhu, R. P. Bora, A. Barman, R. Singh et R. Prabhakar. Dimerization of the full-length alzheimer amyloid beta-peptide (aβ42) in explicit aqueous solution : a molecular dynamics study. *J. Phys. Chem. B*, 116(15): 4405–4416, 2012.

[539] H. Y. Zoghbi et H. T. Orr. Glutamine repeats and neurodegeneration. *Annu. Rev. Neurosci.*, 23:217–247, 2000.

[540] C. Zuccato, M. Valenza et E. Cattaneo. Molecular mechanisms and potential therapeutical targets in huntington's disease. *Physiol. Rev.*, 90:905–981, 2010.

Annexe I

Supporting Information for Chapiter 6

Distinct Dimerization for Various Alloforms of the Amyloid Beta-Protein : Aβ_{1-40}, Aβ_{1-42} and Aβ_{1-40}(D23N)

Sébastien Côté[†], Rozita Laghaei[†], Philippe Derreumaux[‡], and Normand Mousseau[†]

† Département de Physique and Groupe d'étude des protéines membranaires (GÉPROM), Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec), Canada

‡ Laboratoire de Biochimie Théorique, UPR 9080 CNRS, Institut de Biologie Physico Chimique, Institut Universitaire de France, Université Paris Didero - Paris 7, 13 rue Pierre et Marie Curie, 75005 Paris, France

Chemical shifts of Aβ_{1-40} and Aβ_{1-42} monomers

In Figure I.1, we show the chemical shifts of the structural ensembles of Aβ_{1-40} and Aβ_{1-42} monomers obtained at 300K using the OPEP force field coupled to HT-REMD [108]. We computed the chemical shifts using SPARTA+ [429]. Our results agree well to experimental chemical shifts [192] as shown by the Pearson correlation coefficient (PCC). Our method, HT-REMD coupled with the OPEP force field, thus provide consistent results and can be used to simulate the Aβ dimers. Other groups [453, 512] also compared their simulation results on Aβ monomers to these experimental chemical shifts. They obtained similar agreement to us (Table I.I).

Time evolution of A$\beta_{42}\Delta$(41-42) secondary structures

In Figure I.2, we present the evolution of the secondary structure of A$\beta_{42}\Delta$(41-42) starting from the replicas of Aβ_{1-42} at 925 ns.

Solvent accessible surface area

In Figures I.3 and I.4, we analyze the impact of the N-terminal and residues Ile41-Ala42 on the solvent accessible surface area.

Supporting Tables

Table I.I – Pearson correlation coefficient of other computational studies.

Groups	$A\beta_{1-40}$			$A\beta_{1-42}$		
	C_α	C_β	N	C_α	C_β	N
Us	0.9872	0.9972	0.8752	0.9896	0.9978	0.8350
Yang et al. [512]	0.9940	-	0.9420	0.9950	-	0.8830
Takeda et al. [453]	0.9840	0.9995	-	-	-	-

The comparison of previous simulations of the Aβ monomers [108, 453, 512] to the same experimental chemical shifts [192] using Pearson correlation coefficients (PCC).

Supporting Figures

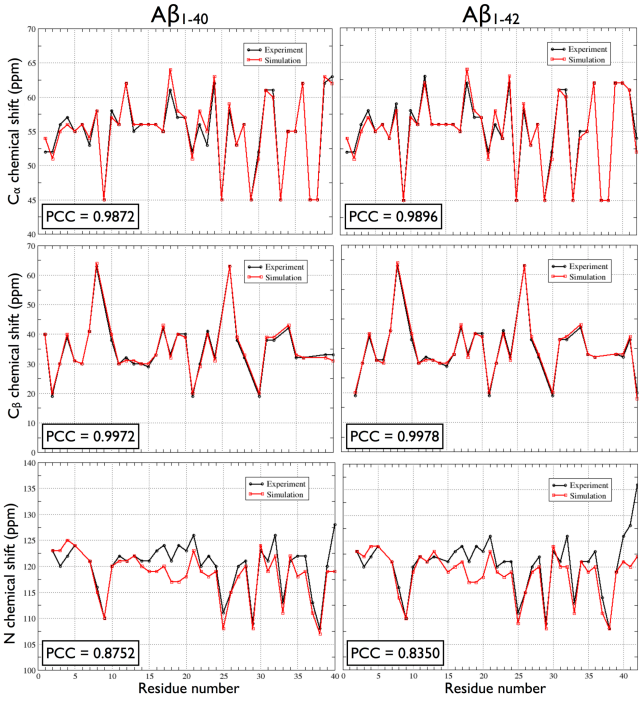


Figure I.1 – Chemical shifts of Aβ_{1-40} and Aβ_{1-42} monomers. The chemical shifts of the structural ensembles of Aβ_{40} and Aβ_{42} monomers at 300K are computed using SPARTA+ [429]. The simulation results on these monomers have been previously analyzed [108]. Our results (red square) are compared with experimental chemical shifts of Hou et al. [192] (black circle). The Pearson correlation coefficients (PCC) indicate agreement with experimental results.

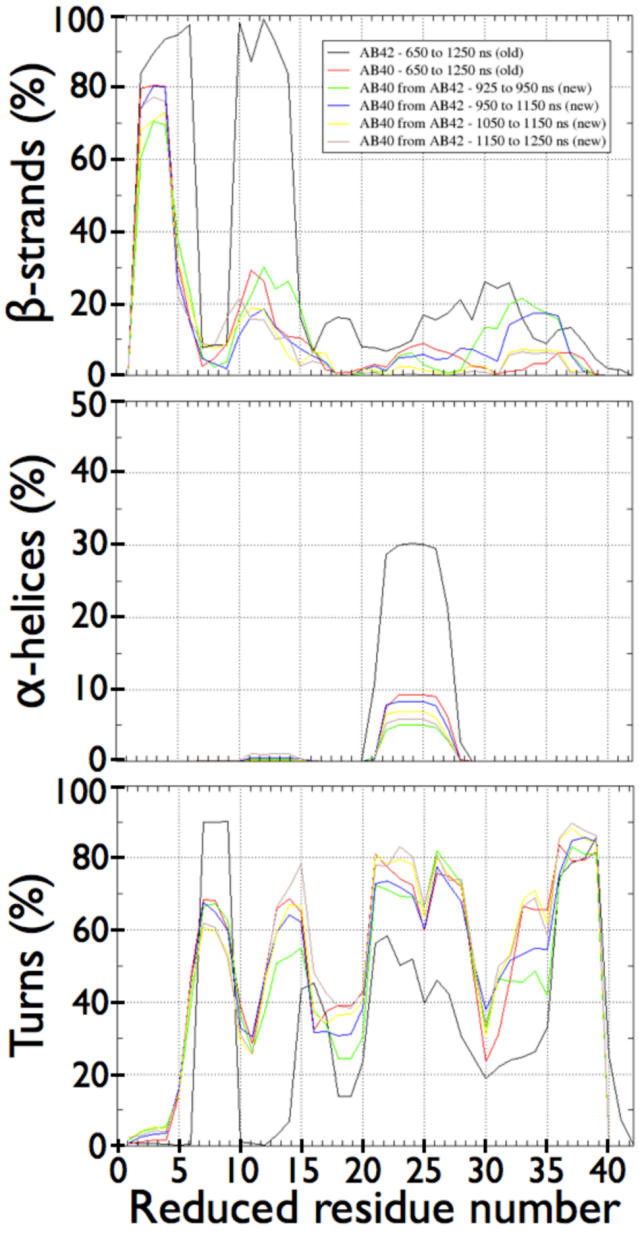


Figure I.2 – Secondary structure of Aβ_{40} from Aβ_{42}'s snapshot at 925 ns. As expected, the values previously obtained for Aβ_{42} (black) relax toward the values previously obtained for Aβ_{40} (red). The initial configuration is the 22+4 replicas of Aβ_{42} without Ile41-Ala42 at 925 ns. Starting from 925 ns, the relaxation is gradual from the 925 to 950 ns interval (green), then 950 to 1050 ns (blue), then 1050 to 1150 ns (yellow), and finally 1150 to 1250 ns (brown). The interval 1050 to 1250 ns (yellow and brown) shows that the secondary structures have relaxed to the values previously obtained for Aβ_{40} (red).

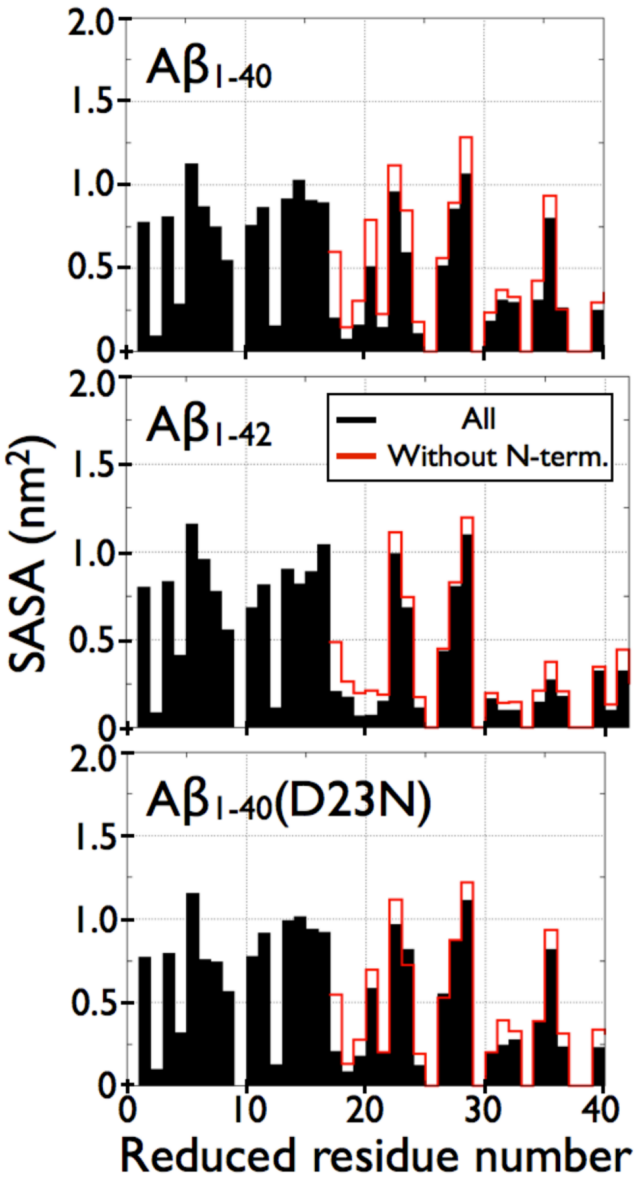


Figure I.3 – Effect of N-terminal on the solvent accessible surface area. For these three alloforms, the truncation of the N-terminal (residues 1-16) increases the solvent accessible surface area (SASA) of the CHC and C-terminal residues as seen by comparing the black curve (considering all residues) to the red curve (without N-terminal).

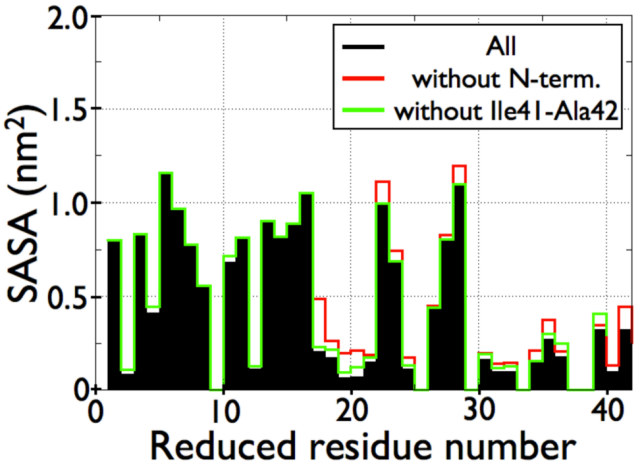


Figure I.4 – Effect of Ile41-Ala42 in Aβ_{42} on the solvent accessible surface area. Truncation of residues Ile41-Ala42 in Aβ_{42} causes a smaller increase of the solvent accessible surface area (SASA) of the CHC and C-terminal residues than the N-terminal truncation (respectively green vs. red).

Annexe II

Supporting Information for Chapiter 9

Free-energy landscape of the amino-terminal fragment of huntingtin in aqueous solution

Vincent Binette$^{\triangle}$, Sébastien Côté$^{\triangle}$, and Normand Mousseau

Département de Physique and Groupe d'étude des protéines membranaires (GÉPROM), Université de Montréal, Montréal (Québec), Canada

$\triangle$ Vincent Binette and Sébastien Côté contributed equally to this work.

HREXMetaD on Htt17_coil

We perform a second HREXMetaD simulation on Htt17 starting from a random coil state (Htt17_coil) to assess the robustness of our simulations starting from the NMR model obtained in the presence of DPC micelles (Htt17_nmr). Both systems have the same size and number of water molecules. Both simulations are run at 303K using 16 scales spanning 1.0 to 0.3 with the same intermediate scales. Exchanges between neighboring scales are attempted every 4 ps resulting in an exchange rate of about 20–40%.

The free energy surfaces (FES) in terms of S_α and S_{gyr} for these two simulations indicate that the main features and the extend of the basin are very similar in both simulation sets (compare Figure II.1 on Htt17_nmr to Figure II.4 on Htt17_coil). The global α-helix probability is also similar : 29.3 $\pm$ 0.7% for Htt17_nmr and 26.9 $\pm$ 0.3% for Htt17_coil. While residues 7 to 12 are less structured in Htt17_coil, the main features of the per residue secondary structure are also preserved : moderate helical content for the first residues, presence of a turn between residues 10 and 13 and more disordered or the last residues (Figure II.9). The FES in terms of the number of helical H-bonds (horizontal axis, S_α) and the gyration radius (vertical axis, S_{rg}) displays a similar conformational ensemble characterized by two main structures : two-helix bundle structures (clusters 1,2,4,5), with the first half of the peptide more structured than the second half (cluster 1,4) and almost fully random structures (clusters 3) as shown in Figure II.9.

Both kinds of structures were also identified in the simulation Htt17_nmr. The FES in terms of the number of helical H-bonds (S_α) and the solvent accessible surface area (SASA) of Htt17's non-polar residues is also similar in both cases as most of the structural ensemble is located between 3 and 5 nm^2 (data not shown). Finally, the contact maps show that the key non-polar and electrostatic contacts are preserved : Met8–Phe17 (24.1% for Htt17_nmr vs. 26.8% for Htt17_coil), Glu5–Lys9 (50.3% for Htt17_nmr vs. 44.5% for Htt17_coil), Glu12–Lys9 (42.3% for Htt17_nmr vs. 38.2% for Htt17_coil), and Glu12–Lys15 (64.0% for Htt17_nmr vs. 58.5 for Htt17_coil) (data not shown).

Overall, we observe an excellent agreement between these two simulations that start from the two very different initial states indicating adequate convergence assessment and sampling of the conformational ensemble.

HREXMetaD vs. PTMetaD for Htt17

In addition to our HREXMetaD simulation on Htt17, we use a second methodology that is very popular – parallel tempering metadynamics (PTMetaD) – to compute the free energy surface of Htt17 in terms of S_α and S_{gyr}. Parallel tempering is often used on its own to simulate protein folding because it increases the probability of escaping free energy minima by allowing exchanges between simultaneous MD simulations at different temperatures [181, 449]. Similarly to HREXMetaD, the combination of MetaD and PT dumbed PTMetaD allows one to correctly sample other CVs not explicitly taken into account by the time-dependent biased potential as demonstrated from proteins with similar conformational ensemble to Htt17 [18, 76, 77]. The temperature distribution used for PT spans 278 to 646K and the intermediate temperatures are determined using a recent protocol and requiring an exchange rate of approximately 20% for a total of 64 replicas [386].

The free energy surface in terms of the number of helical H-bonds (S_α) and gyration radius (S_{gyr}) obtained using PTMetaD is shown in Figure II.10. We observe that its extent is very similar to that obtained using HREXMetaD (Figure II.9), while there are two minor differences : (i) the FES minimum is now bounded between 3 and 6 helical H-bonds – instead of between 2 and 6 for the HREXMetaD simulations – and has narrower gyration radius bracket and (ii) the fully random structures are slightly less favoured when using PTMetaD.

Even if these changes in the FES lead to a slight increase of the α-helical propensity from

29.3 $\pm$ 0.7% (HREXMetaD) to 38$\pm$ 3% (PT-MetaD), the main features of the secondary structure per residue profile are unchanged with the first half of the peptide being more structured than the second half (compare Figures II.9 and II.10). From the same Figures, the cluster analysis of the structures sampled in the FES minima (below 5 kJ/mol) further indicates that our HREXMetaD and PTMetaD simulations sample a similar structural ensemble. Indeed, in agreement with our HREXMetaD simulations, we see that Htt17 mostly adopts a two-helix bundle structure (see clusters 1,2 and 4).

The good agreement between our PTMetaD and HREXMetaD simulations demonstrates the robustness of the sampling in both methodologies although we believe that HREX might escape local minima faster than PT because the configurations have significantly less replicas to diffuse in. The use of HREXMetaD might then reduce the probability that MetaD adds wrong biases to the free energy landscape when compared to PTMetaD.

Comparison to the solution NMR experiment on Htt17

With large intensities for H^{α}(i)-H^{N}(i+1) NOEs, medium intensities for H^{α}(i)-H^{N}(i) NOEs and very small ones for medium range NOEs, the structural ensemble sampled during our simulations is largely compatible with the NMR experiment on Htt17 in aqueous solution indicating that it is mostly unstructured in solution [459] (Figure II.11).

We refine our analysis by comparing the interproton NOEs for our most scaled replica (replica 16) to our unscaled replica (replica 1) that populate a very different conformational ensemble : the helix propensity is only 3.3 $\pm$ 0.1% for the most scaled replica, while it is 36.9 $\pm$ 0.9% for the unscaled replica. We find an almost identical trend for H^{α}(i)-H^{N}(i) interproton distances, slightly weaker H^{N}(i)-H^{N}(i+1) and H^{α}(i)-H^{N}(i+2) intensities, and stronger H^{α}(i)-H^{N}(i+1) intensities (Figure II.11). For its part, the medium-range H^{α}(i)-H^{N}(i+3) NOEs are weaker (data not shown). Taken together, this indicates that the structural ensemble in terms of NOEs of the unscaled replica is dominated by mostly random conformations as the H^{α}(i)-H^{N}(i+1) intensities are very large with a small population of helical conformations as the H^{α}(i)-H^{N}(i+2) and H^{α}(i)-H^{N}(i+3) intensities are stronger compared to the most scaled replica.

We also compute the interproton NOE intensities for two extreme cases : a perfect α-helix and a completely extended conformation. The ideal conformations are build with PYMOL and then minimized with the conjugate gradient method to avoid structural clashes. The results are

shown in Figure II.11. Comparing with both sets, we conclude the intensities observed are consistent with mostly disordered structures.

We note, moreover that we find very low H^N(i)-H^N(i+1) intensities (below 0.05) except for two residues where the intensities drastically increase to 0.7 indicating that only small structural changes can lead to large fluctuations of this NOE.

Overall, our investigation of the interproton distances shows that a globally good agreement between our simulations and NMR experiments. Indeed, three out of four interproton distances are well conserved with large H^α(i)-H^N(i+1), medium H^α(i)-H^N(i) and very small H^α(i)-H^N(i+2) intensities. We find high intensities for the H^N(i)-H^N(i+1) NOEs, which seems conflicting with the NMR experiment showing very weak intensities. The lack of sequential H^N(i)-H^N(i+1) NOE in the NMR experiment is an indicator of a mostly disorder structural ensemble.A more thorough examination of our simulations leads us to believe that the difference is due only to very small and local structural changes as both the fully extended conformation and the structural ensemble sampled by the most scaled replica (having only $3.3 \pm 0.1\%$ of helical content) have H^N(i)-H^N(i+1) intensities of 0.7 and higher. The presence of H^α(i)-H^N(i+3) NOEs indicates a small population of helical structures, not found in NMR experiments. We thus conclude that structural ensemble of Htt17 is, at the exception of very local flexibility and small overestimation of the helical content, in agreement with this experiment.

Supporting Figures

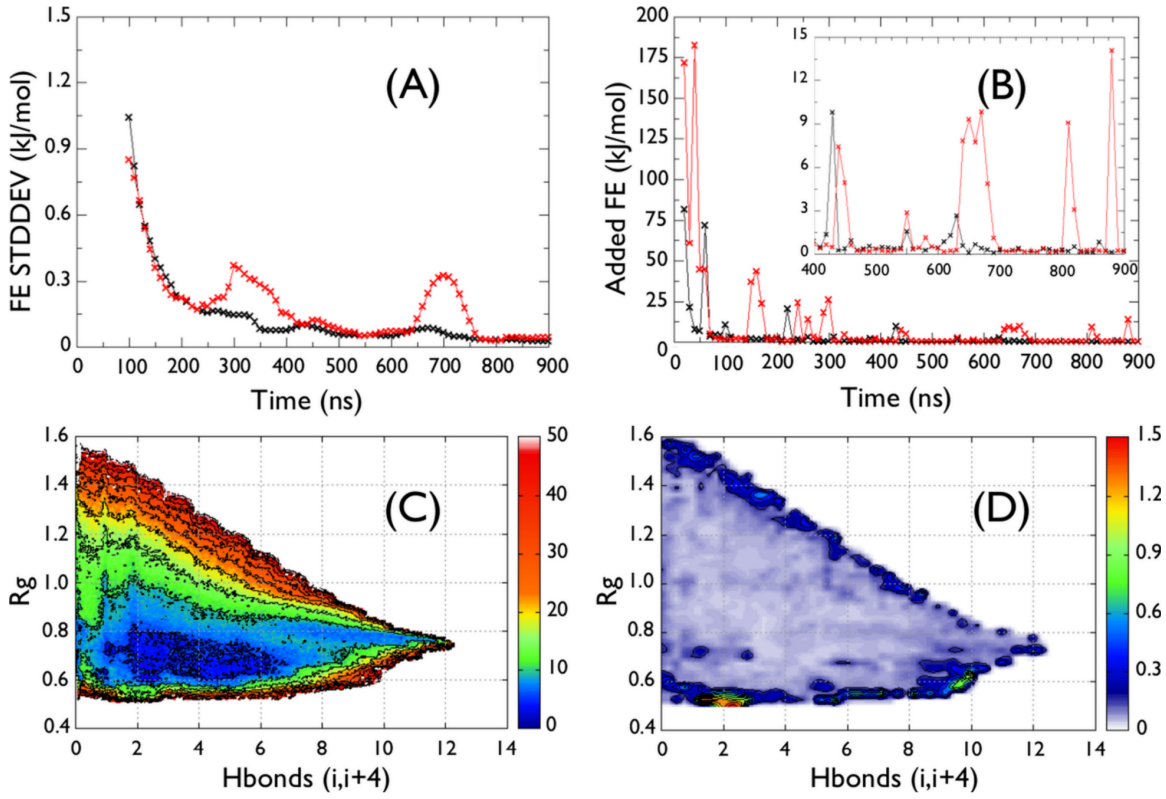


Figure II.1 – Convergence assessment of the Htt17_nmr simulation. **(A)** Running average of the standard deviation of the 1D-FES (S_α in black and S_{gyr} in red) over 100 ns time-windows. **(B)** Total addition of free energy to the FES every 10 ns. **(C)** The 2D-FES (S_α ;S_{gyr}) and **(D)** its uncertainty computed on the convergence interval (400–900 ns), which is determined from the small modifications of the FES after 400 ns shown in (A) and (B). We observe that the uncertainty on the FES is mostly located to its border, while it is low ($<$ 0.5 kJ/mol) inside the basin. Energy isolines are drawn every 5 kJ/mol for **(C)** and 0.15 kJ/mol and **(D)**.

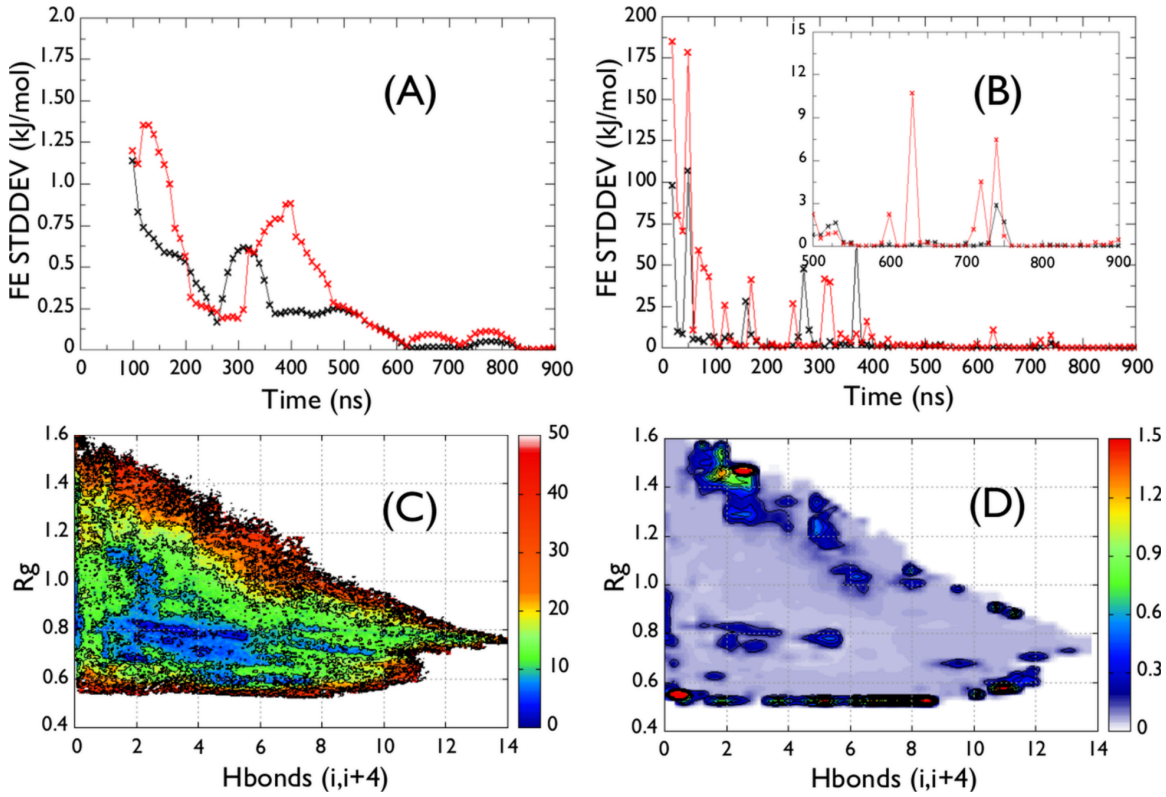


Figure II.2 – Convergence assessment of the Htt17Q$_{17}$ simulation. **(A)** Running average of the standard deviation of the 1D-FES (S_α in black and S_{gyr} in red) over 100 ns time-windows. **(B)** Total addition of free energy to the FES every 10 ns. **(C)** The 2D-FES (S_α ;S_{gyr}) and **(D)** its uncertainty computed on the convergence interval (500–900 ns), which is determined from the small modifications of the FES after 500 ns shown in (A) and (B). We observe that the uncertainty on the FES is mostly located to its border, while it is low ($<$ 1.0 kJ/mol) inside the basin. Energy isolines are drawn every 5 kJ/mol for **(C)** and 0.15 kJ/mol and **(D)**.

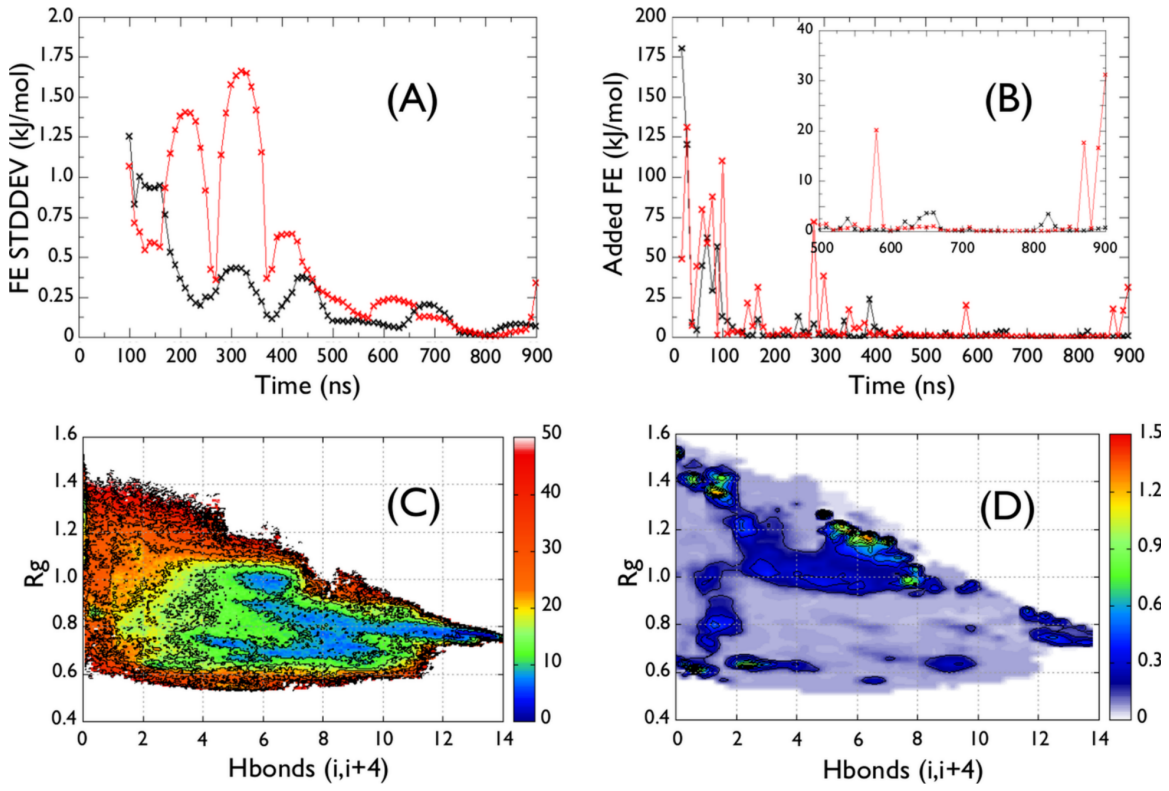


Figure II.3 – Convergence assessment of the Htt17$Q_{17}P_{11}$ simulation. **(A)** Running average of the standard deviation of the 1D-FES (S_α in black and S_{gyr} in red) over 100 ns time-windows. **(B)** Total addition of free energy to the FES every 10 ns. **(C)** The 2D-FES (S_α ;S_{gyr}) and **(D)** its uncertainty computed on the convergence interval (500–900 ns), which is determined from the small modifications of the FES after 400 ns shown in (A) and (B). We observe that the uncertainty on the FES is mostly located to its border, while it is low ($<$ 1.0 kJ/mol) inside the basin. Energy isolines are drawn every 5 kJ/mol for **(C)** and 0.15 kJ/mol and **(D)**.

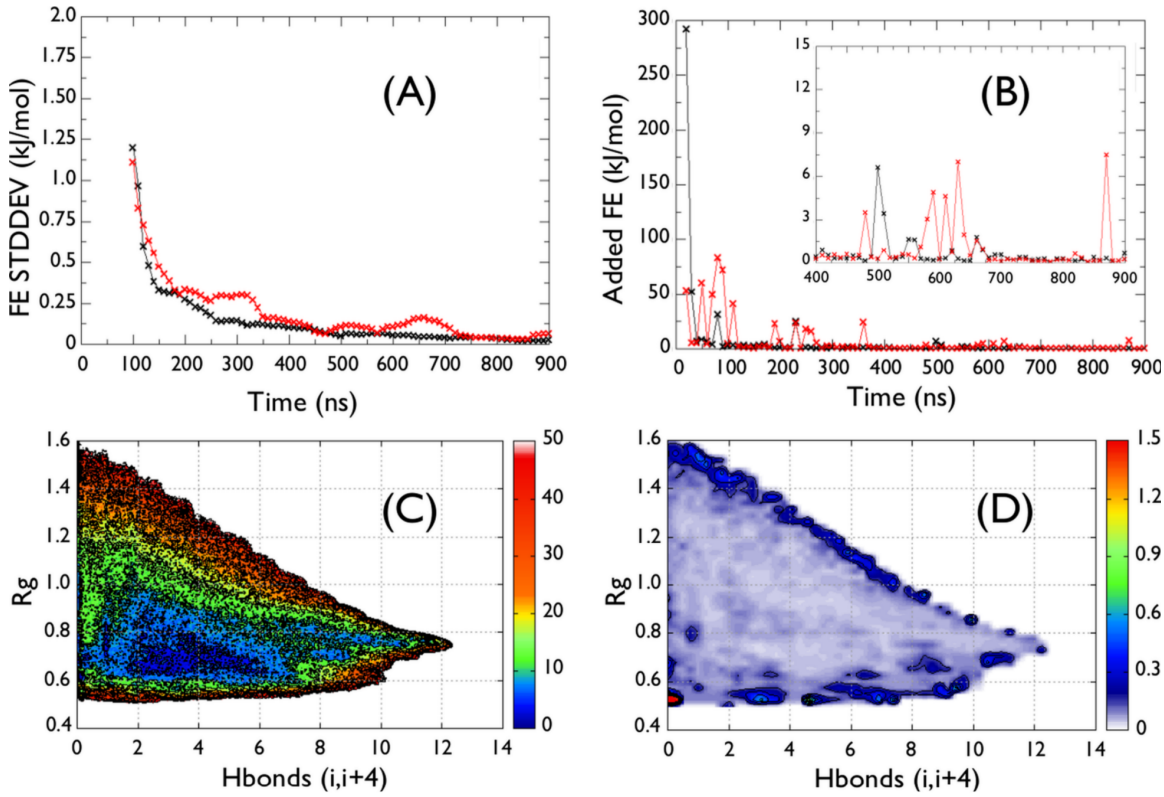


Figure II.4 – Convergence assessment of the Htt17_coil simulation. **(A)** Running average of the standard deviation of the 1D-FES (S_α in black and S_{gyr} in red) over 100 ns time-windows. **(B)** Total addition of free energy to the FES every 10 ns. **(C)** The 2D-FES (S_α ;S_{gyr}) and **(D)** its uncertainty computed on the convergence interval (400–900 ns), which is determined from the small modifications of the FES after 400 ns shown in (A) and (B). We observe that the uncertainty on the FES is mostly located to its border, while it is low ($<$ 0.5 kJ/mol) inside the basin. Energy isolines are drawn every 5 kJ/mol for **(C)** and 0.15 kJ/mol and **(D)**.

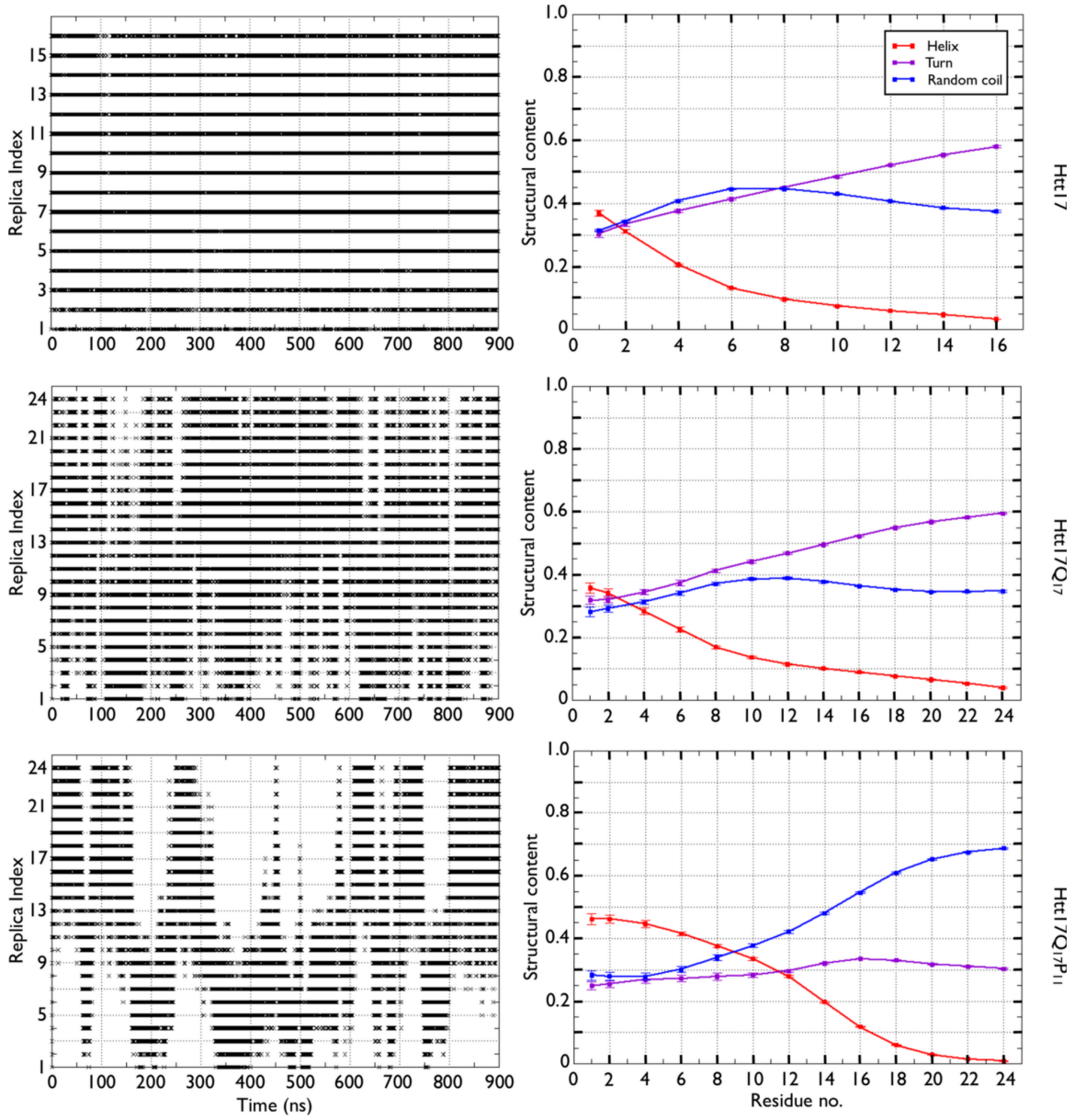


Figure II.5 – Sampling assessment of the HREX simulations. The left panel shows the replica index visiting the first scale. The right panel shows the secondary structure as a function of the scaling. Htt17Q_{17} and Htt17$Q_{17}P_{11}$ are respectively shown from top to bottom.

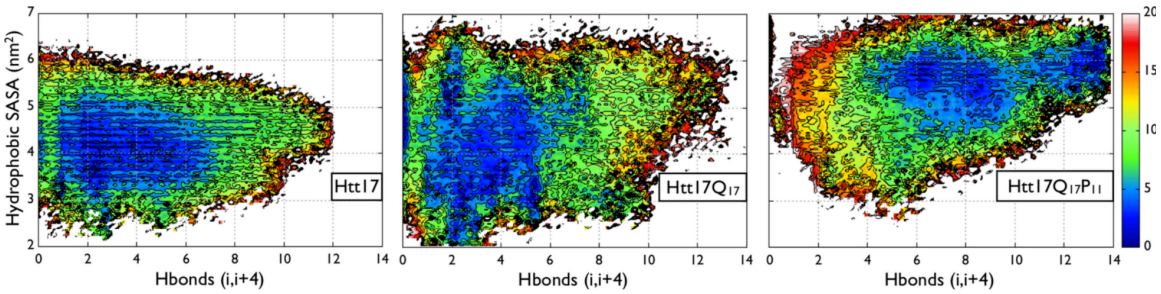


Figure II.6 – The FES of the Htt17 segment as a function of the number of helical H-bonds (S_α, horizontal axis) and SASA of Htt17's non-polar residues (vertical axis) is shown for Htt17, Htt17Q_{17} and Htt17$Q_{17}P_{11}$ from left to right. Energy isolines are drawn every 2 kJ/mol.

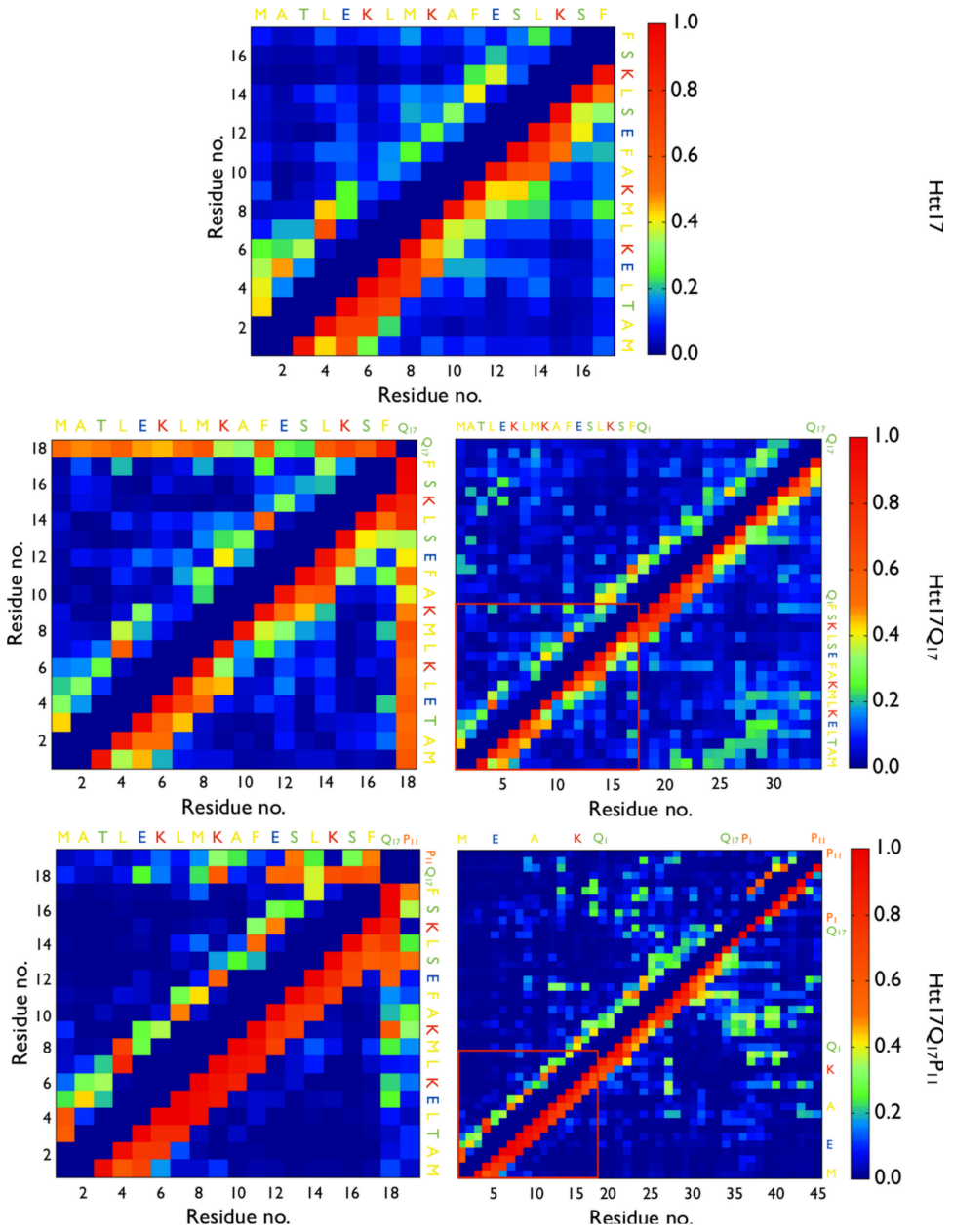


Figure II.7 – Contact maps of Htt17_nmr, Htt17Q_{17} and Htt17$Q_{17}P_{11}$ are shown from top to bottom. The side-chain/side-chain and the total number of contacts are respectively displayed on the upper and lower halves of the contact maps. For Htt17Q_{17} and Htt17$Q_{17}P_{11}$, the global propensity of Q_{17}/Htt17 and P_{11}/Htt17 contacts (left column) and the per residue probability of each individual glutamines and prolines (right column) are shown. The red square indicates the Htt17/Htt17 contacts when appropriate.

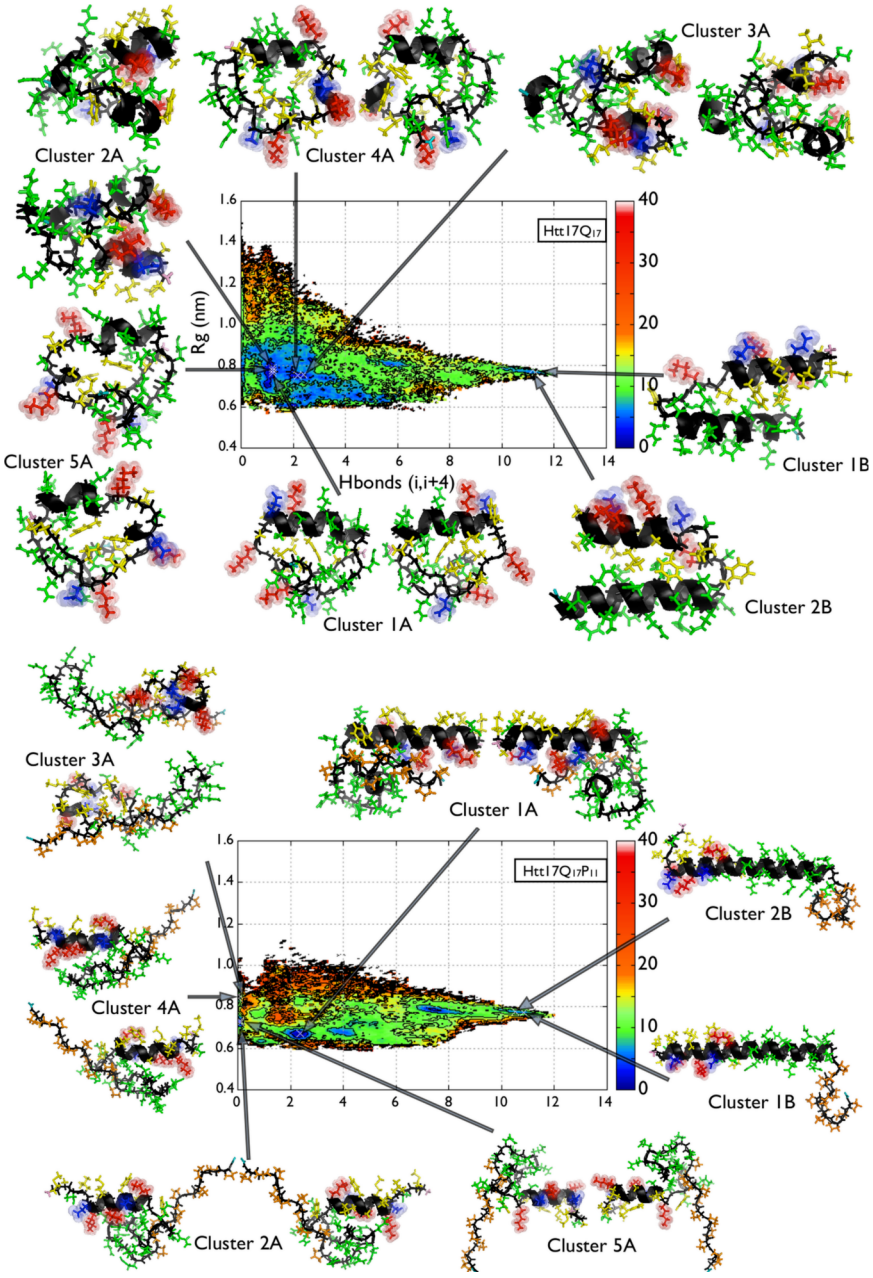


Figure II.8 – The FES of the Q_{17} segment as a function of the number of helical H-bonds (S_α, horizontal axis) and gyration radius (S_{gyr}, vertical axis) is shown for Htt17Q_{17} (top) and Htt17$Q_{17}P_{11}$ (bottom). The main clusters of the conformations inside the main basin (below 4 kJ/mol) and those with more than 9.5 helical H-bonds (below 8 kJ/mol) are displayed around the FES. Energy isolines are drawn every 4 kJ/mol.

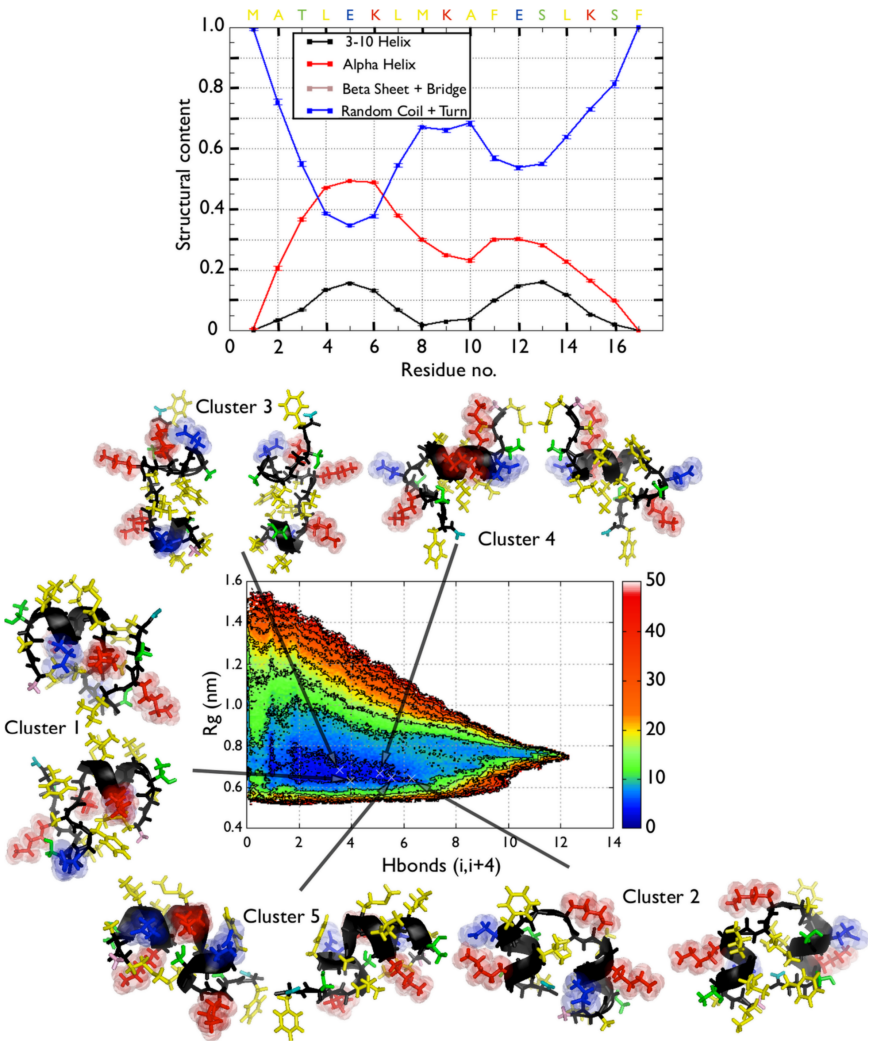


Figure II.9 – The per residue secondary structure of Htt17 from the HREXMetaD simulation starting from the random structure is shown in the top panel. The probability of α-helix, 3-10 helix, β-bridge and β-strand, and all other motifs are respectively shown in red, black, brown and blue. The FES of the Htt17 segment as a function of the number of helical H-bonds (S_α, horizontal axis) and gyration radius (S_{gyr}, vertical axis) is shown in the bottom panel. Energy isolines are drawn every 5 kJ/mol. The FES is surrounded by the cluster center of the representative structures found below 5 kJ/mol.

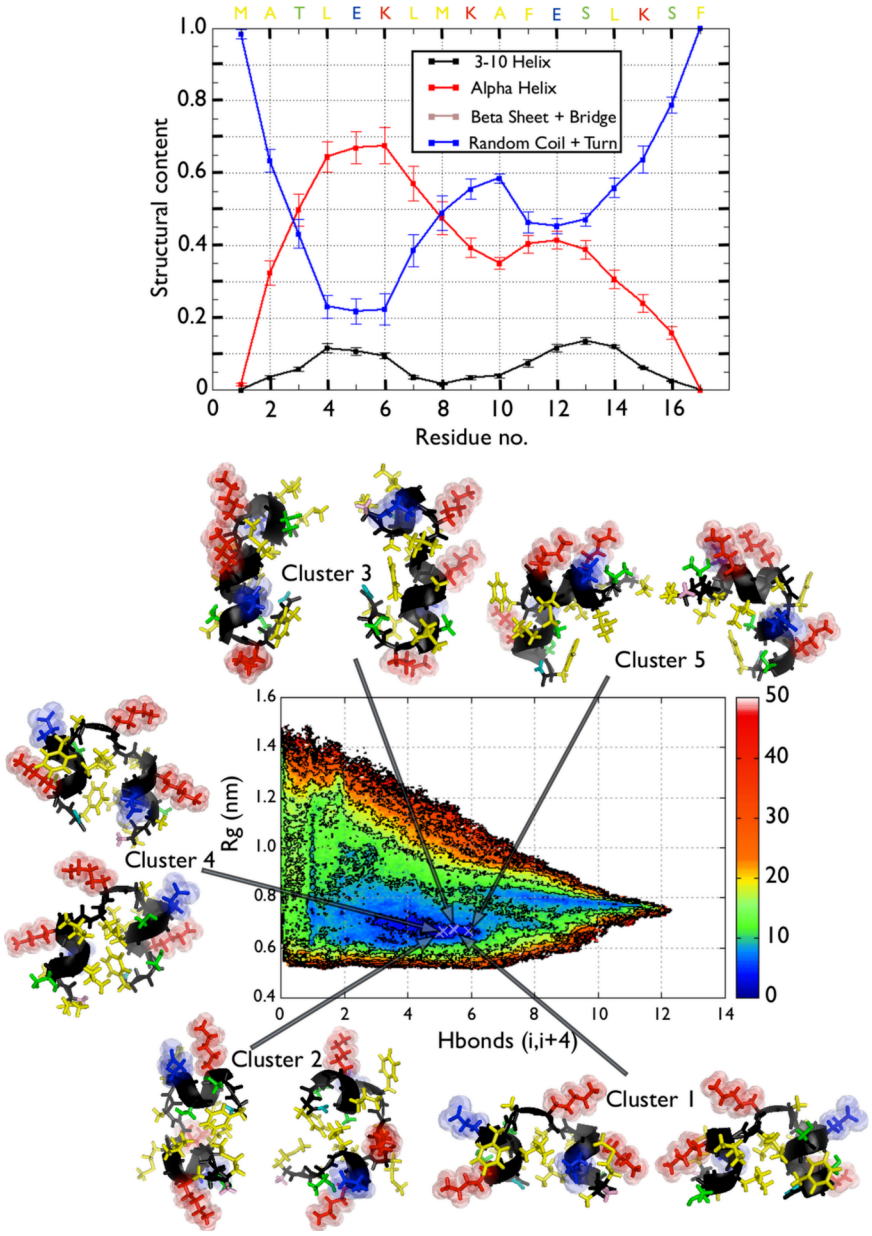


Figure II.10 – The per residue secondary structure of Htt17 from the PTMetaD simulation starting from a random coil structure is shown in the top panel. The probability of α-helix, 3-10 helix, β-bridge and β-strand, and all other motifs are respectively shown in red, black, brown and blue. The FES of the Htt17 segment as a function of the number of helical H-bonds (S_α, horizontal axis) and gyration radius (S_{gyr}, vertical axis) is shown in the bottom panel. Energy isolines are drawn every 5 kJ/mol. The FES is surrounded by the cluster center of the representative structures found below 5 kJ/mol.

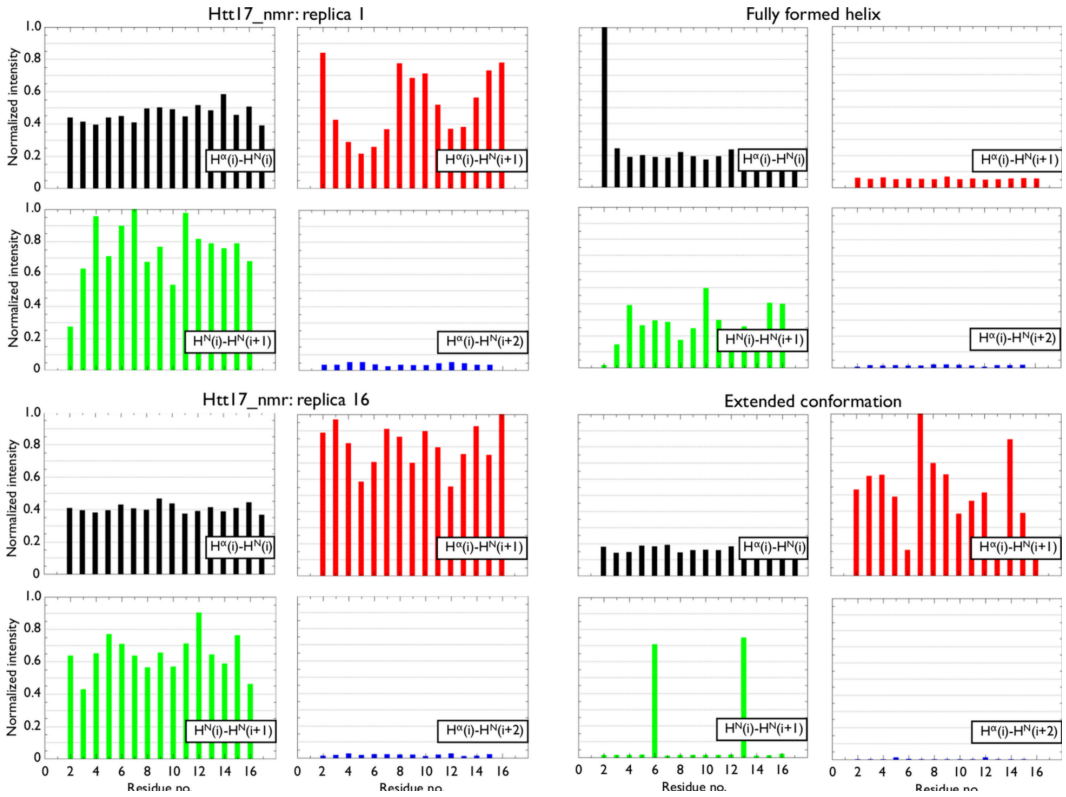


Figure II.11 – The computed intensities of the interproton NOEs for all residues between the H^{α} of residue i and the H^{N} of residues i, i+1 and i+2, as well as between the H^{N} of residues i and i+1. The top left panel shows the NOEs for the analysis replica, the bottom left panel shows the NOEs for the most scaled replica, the top right panel shows the NOEs of a fully formed α-helix and the bottom right panel shows the NOEs of a fully extended conformation.

Annexe III

Supporting Information for Chapiter 10

All-Atom Stability and Oligomerization Simulations of Polyglutamine Nanotubes with and without the 17-Amino-Acid N-terminal Fragment of the Huntingtin Protein

Sébastien Côté[†], Guanghong Wei[‡], and Normand Mousseau[†]

† Département de Physique and Groupe d'étude des protéines membranaires (GÉPROM), Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec), Canada

‡ State Key Laboratory of Surface Physics and Department of Physics, Fudan University, 220 Handan Road, Shanghai, 200433, China

Supporting Figures

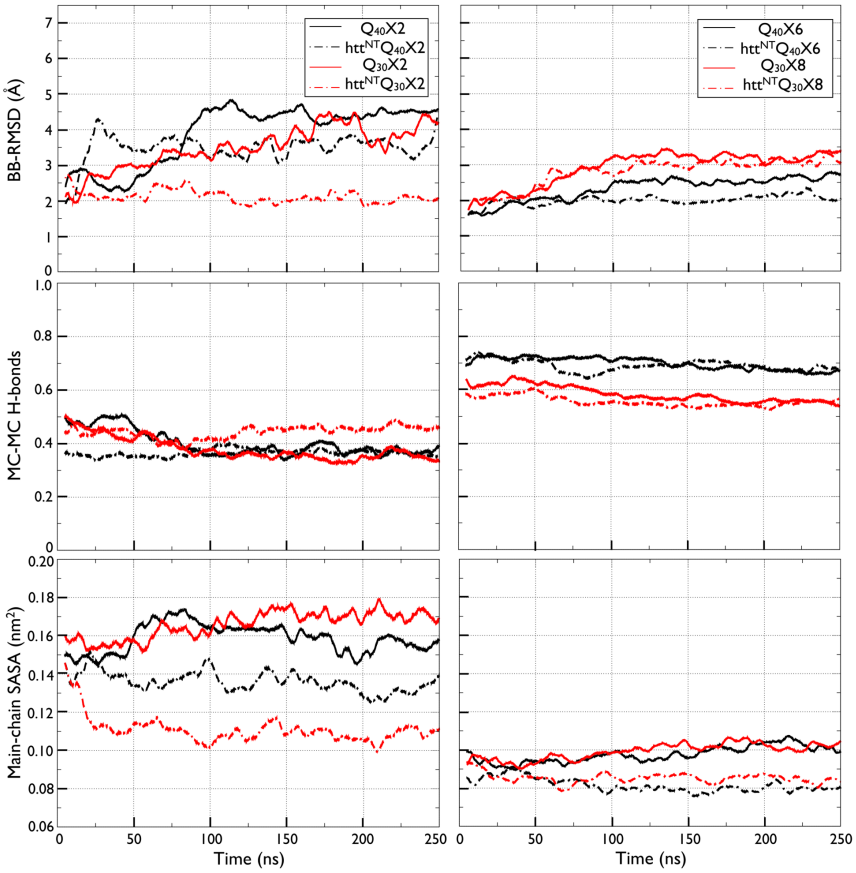


Figure III.1 – Time evolution of various structural properties of the glutamine residues for the dimeric nanotubes (left column) and the longer nanotubes (right column). From top to bottom : the backbone-RMSD (BB-RMSD) measured from the initial structure, the β-sheet propensity and the main-chain solvent accessible area per residue. The values for Q_{40} and Q_{30} are respectively displayed in black and red. The dotted lines represent the simulations with the htt^{NT}. All these quantities show small variations over the last 100 ns, which is considered converged. Data are computed as a running-time average over 5-ns windows.

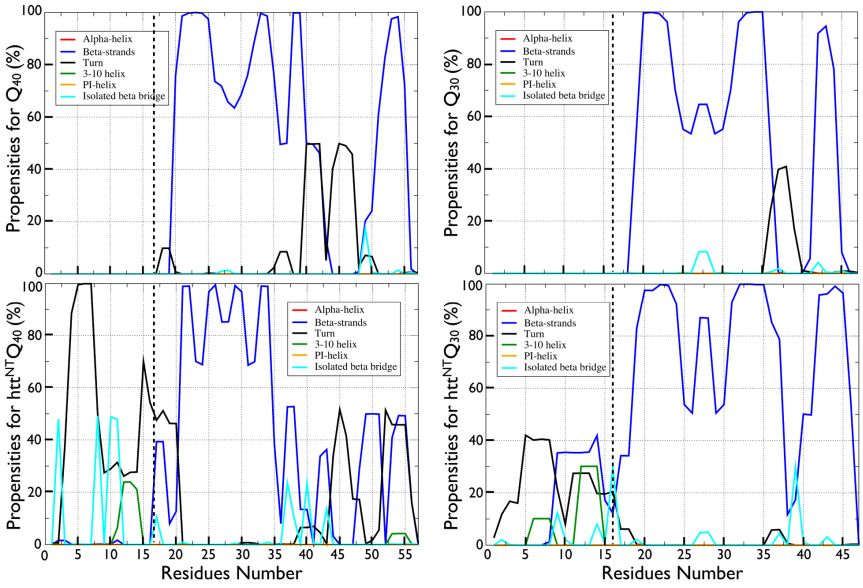


Figure III.2 – Per residue secondary structures of the dimeric nanotubes. To the left, secondary structures for $Q_{40}X2$ and $htt^{NT}Q_{40}X2$. To the right, secondary structures for $Q_{30}X2$ and $htt^{NT}Q_{30}X2$. The thick black dotted line in the graphs of $htt^{NT}Q_{40}$ and $htt^{NT}Q_{30}$ represents residue Phe17, the last residue of the htt^{NT}. Propensities for Q_{40} and Q_{30} have been shifted by 17 residues to ease comparison against $htt^{NT}Q_{40}$ and $htt^{NT}Q_{30}$, respectively. The secondary structure elements are represented by the following colours : α-helix (red), β-strand (blue), turn (black), 3-10 helix (green), π-helix (yellow) and isolated β-bridges (cyan).

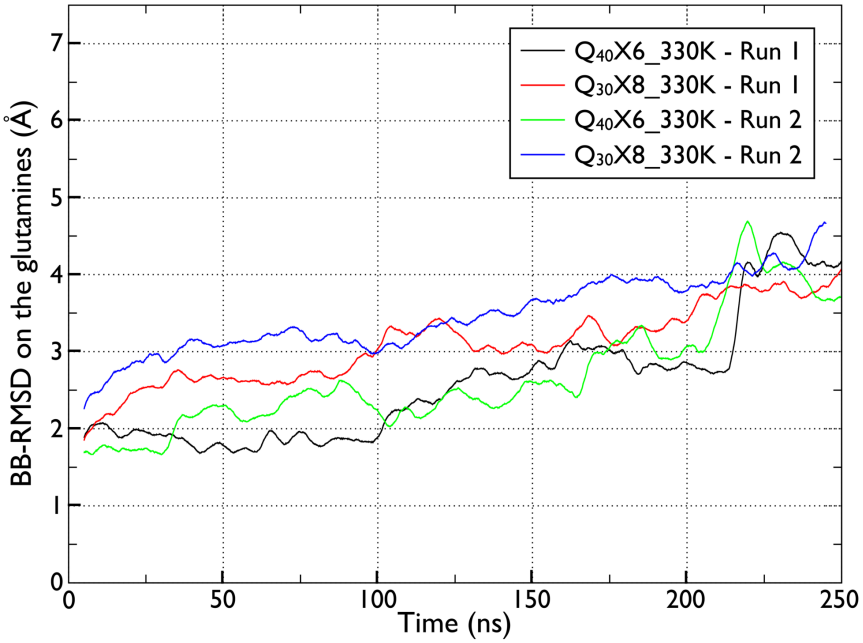


Figure III.3 – Time evolution at 330 K of the backbone root mean square deviation (BB-RMSD) on the glutamine residues as measured from the initial structure. The BB-RMSD is computed on the backbone atoms : O, N, C_α and C. The values for $Q_{40}X6$ are shown in black (run 1) and in green (run 2). The values for $Q_{30}X8$ are shown in red (run 1) and in blue (run 2). Data are computed as a running-time average over 5-ns windows.

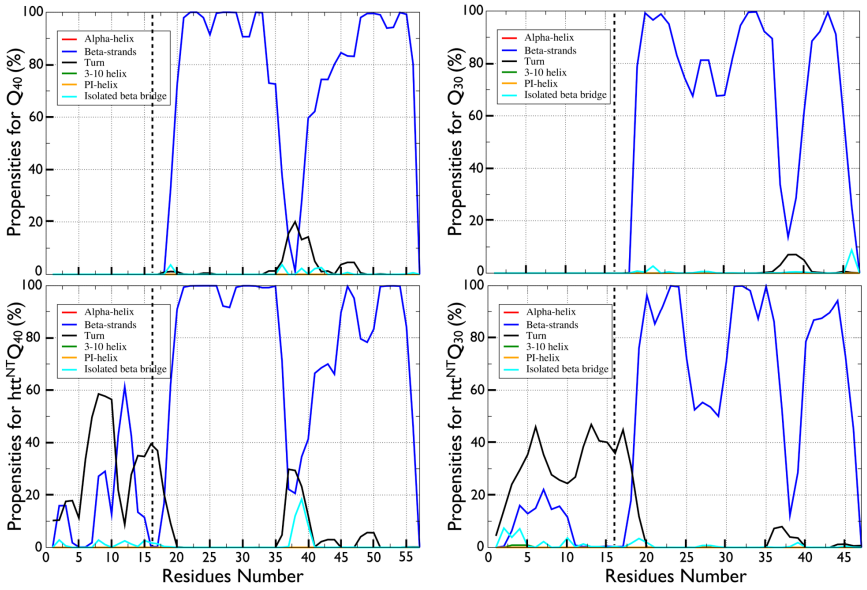


Figure III.4 – Per residue secondary structures of the longer nanotubes. To the left, secondary structures for $Q_{40}X6$ and $htt^{NT}Q_{40}X6$. To the right, secondary structures for $Q_{30}X8$ and $htt^{NT}Q_{30}X8$. The thick black dotted line in the graphs of $htt^{NT}Q_{40}$ and $htt^{NT}Q_{30}$ represents residue Phe17, the last residue of the htt^{NT}. Propensities for Q_{40} and Q_{30} have been shifted by 17 residues to ease comparison against $htt^{NT}Q_{40}$ and $htt^{NT}Q_{30}$, respectively. The secondary structure elements are represented by the following colours : α-helix (red), β-strand (blue), turn (black), 3-10 helix (green), π-helix (yellow) and isolated β-bridges (cyan).

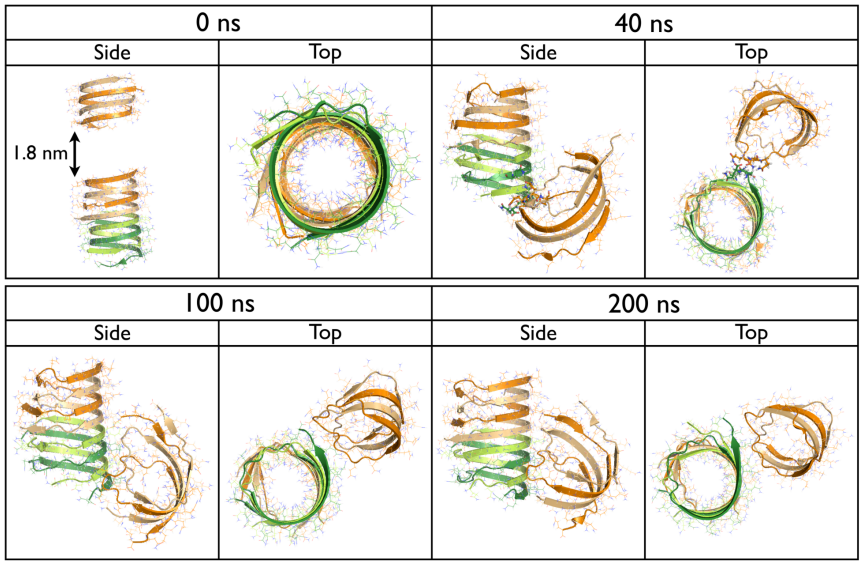


Figure III.5 – Snapshots during the $Q_{40}X6\_Oligo2$ simulation. The side and top views at 0, 40, 100 and 200 ns during the oligomerization between a dimeric and a tetrameric nanotube. Initially, the two nanotubes are positioned at a minimum distance of 1.8 nm. At 40 ns, the dimer binds to the tetramer. The dimeric nanotube then stays on the side of the tetrameric nanotube through the remaining of the simulation.

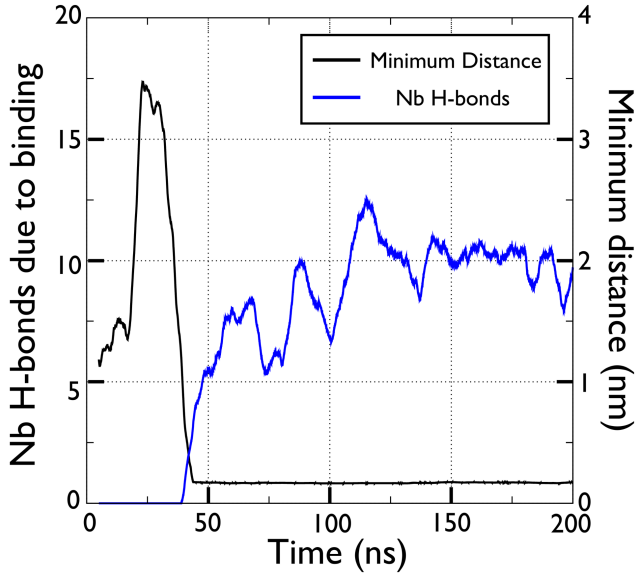


Figure III.6 – Time evolution of the number of hydrogen bonds between the dimeric and the tetrameric nanotube during the $Q_{40}X6\_Oligo2$ simulation. The number of hydrogen bonds (left axis, blue) is correlated to the minimal distance between the dimer and the tetramer (right axis, black).

Annexe IV

Supporting Information for Chapiter 11

Atomistic mechanisms of huntingtin N-terminal fragment insertion on a phospholipid bilayer revealed by molecular dynamics simulations

Sébastien Côté†, Guanghong Wei‡, and Normand Mousseau†

† Département de Physique and Groupe d'étude des protéines membranaires (GÉPROM), Université de Montréal, Montréal (Québec), Canada

‡ State Key Laboratory of Surface Physics and Department of Physics, Fudan University, Shanghai, 200433, China

Choice of forcefield

The htt^{NT} region is known to be an important modulator of the aggregation of the Q_N region in solution [396, 459]. As such, structural results on htt^{NT} could provide a better understanding of its role during aggregation. However, there is no high-resolution experimental structure of this fragment in solution. Most structural information is limited to circular dichroism (CD) experiments showing that the equilibrium structural ensemble of htt^{NT} populates between 10 to 55% α-helical conformations [16, 319, 459, 504]. Two-dimensional proton TOCSY and NOESY NMR measurements, for their part, suggest that the htt^{NT} peptide has no stable α-helix in solution [459].

It is then crucial to determine a forcefield that is suited for modelling htt^{NT} before focusing on its membrane interactions. To do so, we investigate the stability of α-helical conformations in htt^{NT} in solution using four different forcefields – AMBER99sb*-ILDN, CHARMM27, AMBER03 and OPLS-AA\L – starting from the same initial structure in which the htt^{NT} is a single α-helix. For each forcefield, two independent simulations of at least 500 ns were done starting from the same initial state with different velocity distributions.

From the α-helix propensity as a function of time (Figure IV.3), we note that CHARMM27 clearly overestimates the α-helical propensity since the htt^{NT} peptide stays as a single helix

throughout both simulations. On the other hand, OPLS-AA\L clearly underestimates the α-helical propensity as the helix rapidly unfolds after few nanoseconds and it remains mostly unfolded throughout the remaining of both simulations. The situation is different for AMBER99sb*-ILDN and AMBER03 as they exhibit large fluctuations in the α-helical content in both runs. More precisely, the trajectories obtained using AMBER99sb*-ILDN fluctuate mostly between 0 and 45%, while those obtained using AMBER03 mostly fluctuate between 20 and 85%. Overall, the average α-helix propensity over the last 400 ns for each simulation separately is $1 \pm 4\%$ and $5 \pm 11\%$ (OPLS-AA), $10 \pm 18\%$ and $21 \pm 21\%$ (AMBER99sb*-ILDN), $55 \pm 25\%$ and $79 \pm 12\%$ (AMBER03), and $76 \pm 4\%$ and $82 \pm 8\%$ (CHARMM27).

When compared to the percentages obtained by the four independent CD experiments (10% [319], 34% [504], 45% [16] and 55% [459]) and when considering that NMR chemical shifts suggest that the htt^{NT} peptide has no stable α-helix [459] in solution, we see that AMBER99sb*-ILDN gives the sampling agreeing best with experiments, while AMBER03 gives a distribution of structures shifted towards higher α-helical percentages. This tendency is also clear from an analysis of the conformational clusters populated during each trajectory (Figure IV.4) since the most populated cluster for AMBER03 is a single α-helix sampled 39% and 82% of the time for each simulation respectively. The simulations using AMBER99sb*-ILDN, on the other hand, sample mostly coil–α-helix configurations with the single α-helix being occurring only 2% of the time during the first trajectory. We conclude from this comparative study that AMBER99sb*-ILDN provides the least bias toward α-helical structures for htt^{NT} compared to experiments in solution. We therefore use this forcefield for our investigation on the interaction of htt^{NT}Q$_N$ with lipid membranes.

Supporting Figures

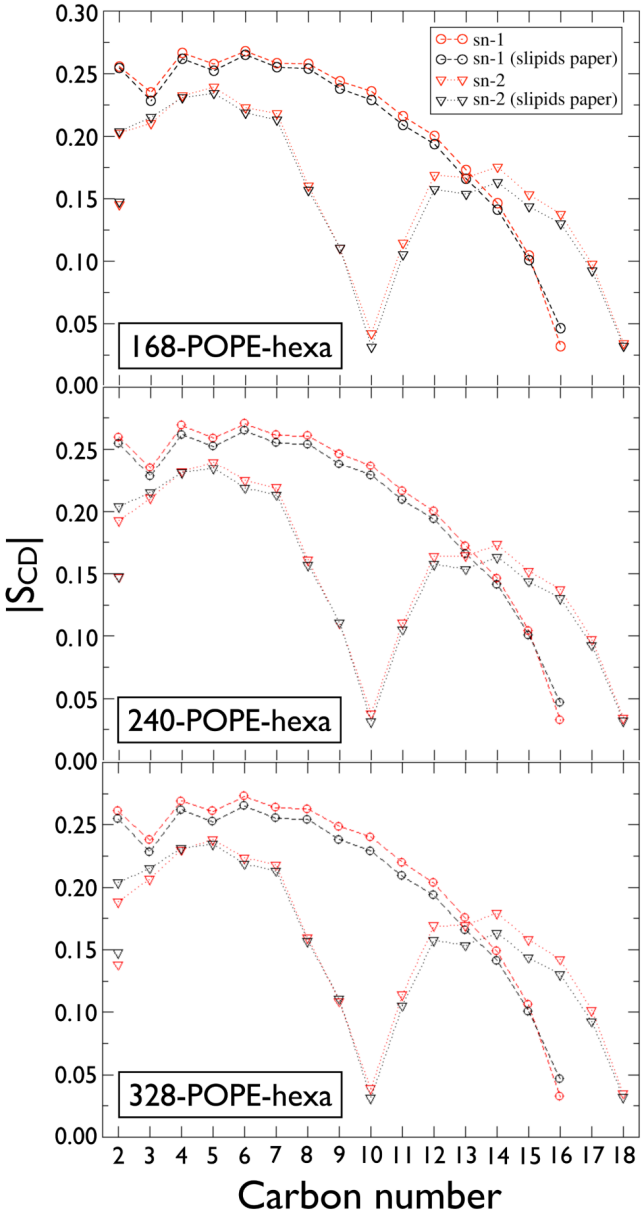


Figure IV.1 – Order parameters of the acyl chains for the three bilayer systems : 168 POPE for htt^{NT}, 240 POPE for htt^{NT}Q$_{10}$ and 328 POPE for htt^{NT}Q$_{20}$. Our values, computed on the 250-350 ns time interval (red curves), are compared to the values computed in the original SLIPIDS paper (black curves) that agreed well with experimental measurements. The order parameters for the palmitic and oleic chains are respectively depicted by empty circles and triangles.

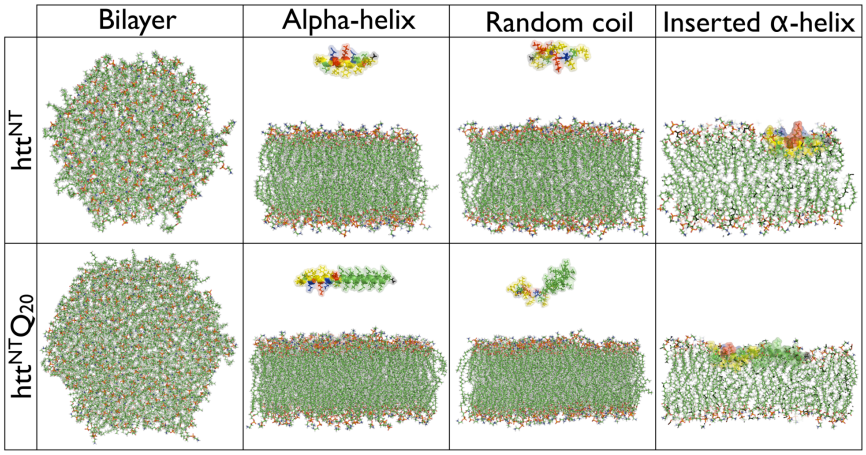


Figure IV.2 – A top view of the phospholipid bilayer used for each type of simulations (first column) and the initial configurations used for the peptide-membrane simulations of htt^{NT} and $htt^{NT}Q_{20}$: a single α-helix (second column) and a random coil (third column) at a minimal distance greater than the real space cutoffs from the membrane, and a single α-helix (last column) inserted in the bilayer.

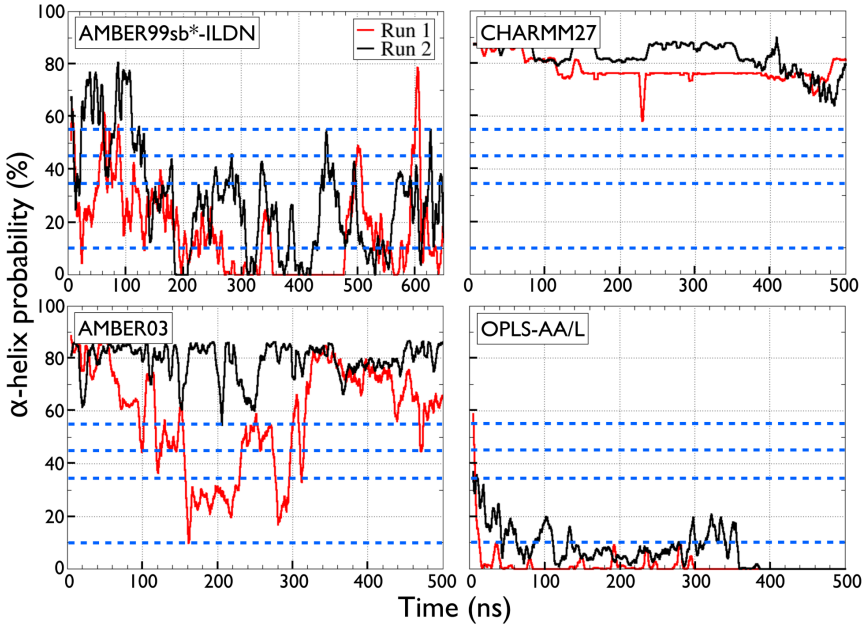


Figure IV.3 – α-helix propensity as a function of time for htt^{NT} in solution. The running time averages taken over 5-ns windows for the first (red) and second (black) simulations are displayed for the four different forcefields used : AMBER99sb*-ILDN, CHARMM27, AMBER03 and OPLS-AA\L. The four CD measurements (10% [319], 34% [504], 45% [16] and 55% [459]) are displayed as blue dotted lines.

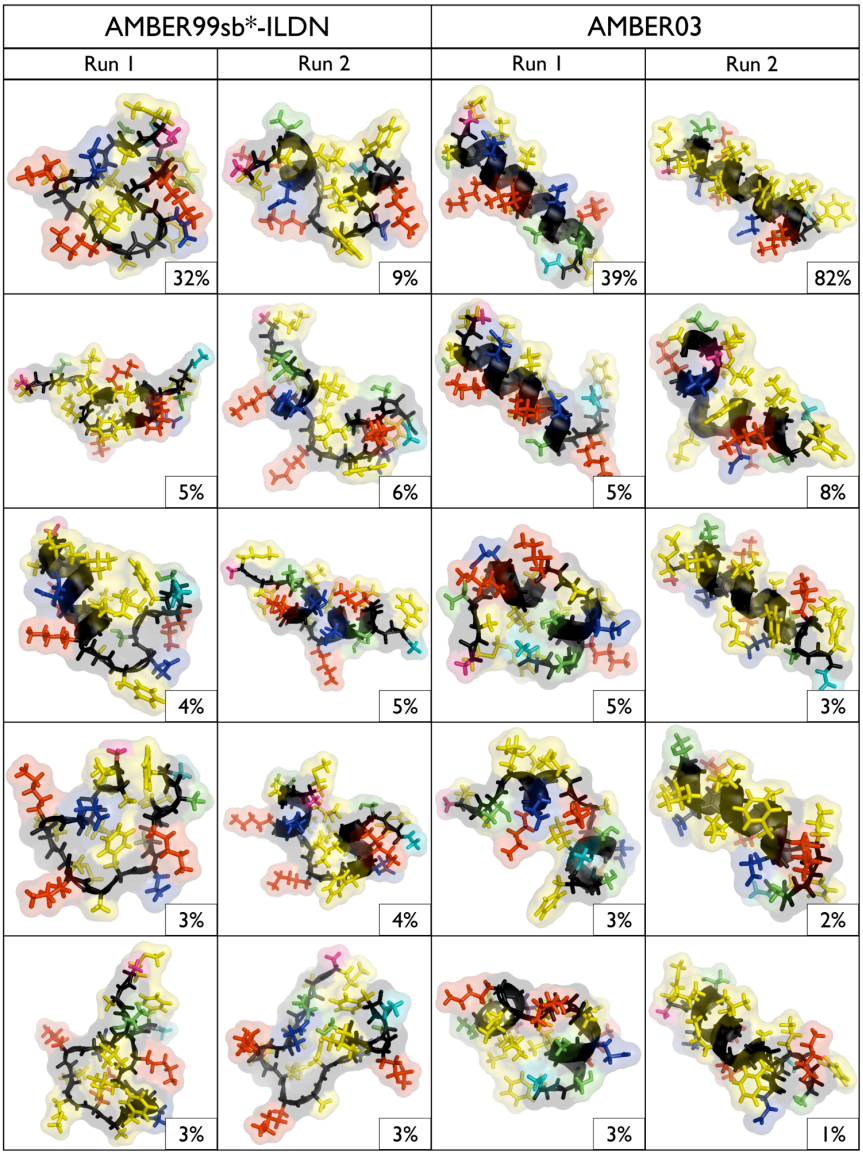


Figure IV.4 – The five main conformation clusters sampled by the htt^{NT} in solution using AMBER99sb*-ILDN and AMBER03 forcefields. The threshold for clustering is a backbone RMSD of 2 Å.

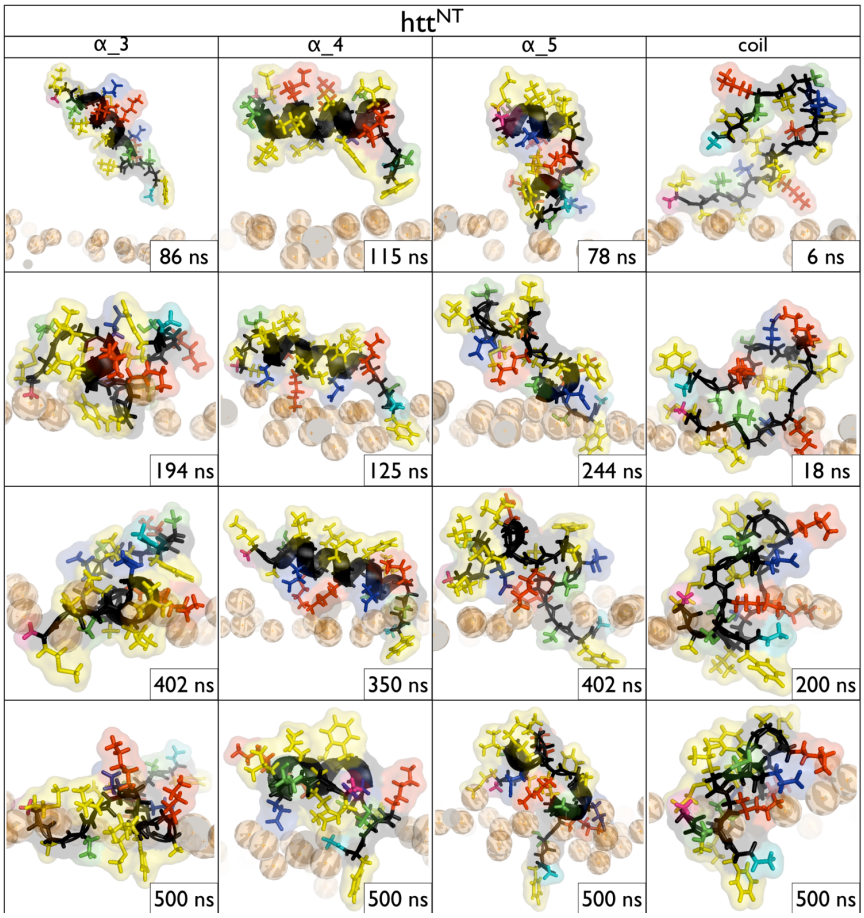


Figure IV.5 – The httNT_α_3 (first column), httNT_α_4 (second column), httNT_α_5 (third column) and httNT_coil (last column) trajectories of htt^{NT}. The non-polar, negatively charged, positively charged and polar amino acids of the peptide are respectively shown in yellow, red, blue and green. The backbone atoms are displayed in black, and the N- and C-terminal are respectively shown in pink and teal. Only the phosphorus atom is displayed as an orange sphere for each phospholipid.

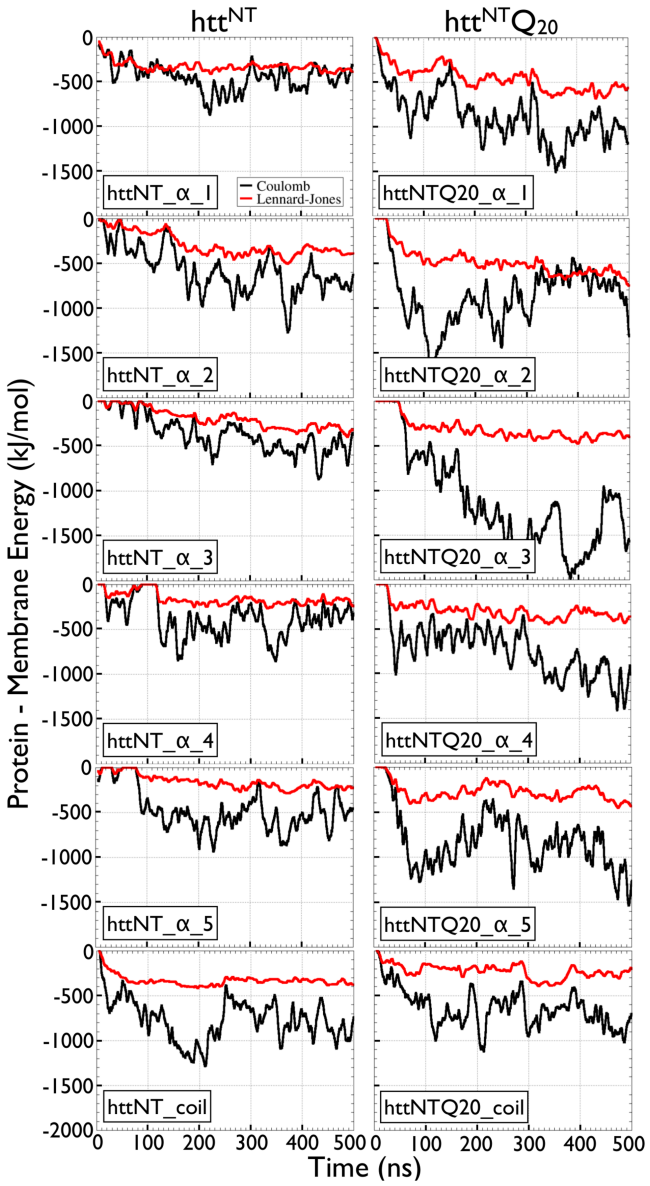


Figure IV.6 – Energies of $htt^{NT}Q_N$ – membrane interactions. The short range Coulomb (i.e. without long range PME) and Lennard-Jones energies are respectively shown in black and red for the htt^{NT} (left column) and $htt^{NT}Q_{20}$ (right column) peptides. The data depicted is the running time averages taken over 5-ns windows.

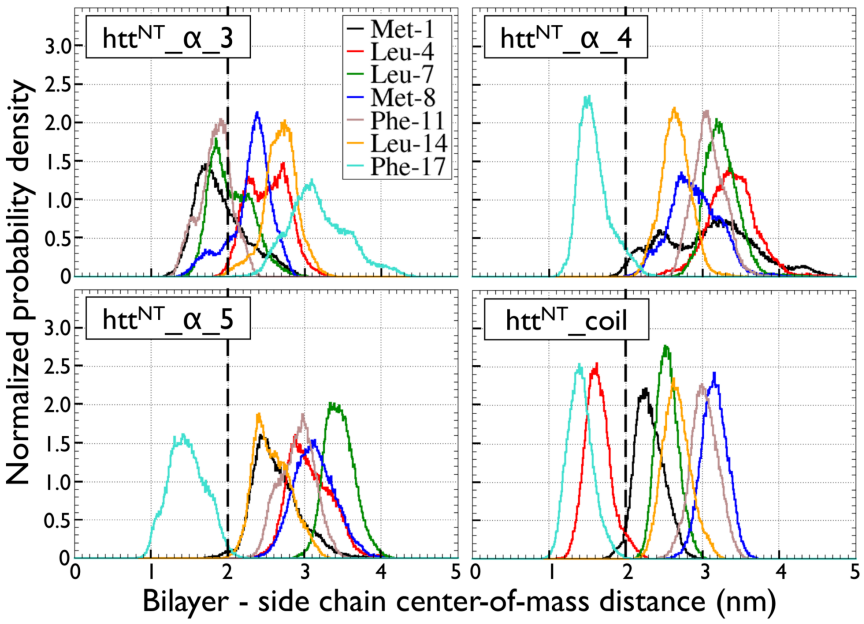


Figure IV.7 – Position probability densities of all non-polar side-chains with respect to the center-of-mass of the phospholipid bilayer of htt^{NT} for httNT_α_3, httNT_α_4, httNT_α_5 and httNT_coil. The average position of the phospholipids' phosphate group is represented by the thick black dotted line near 2.0 Å. The average is taken over the 300-500 ns time interval.

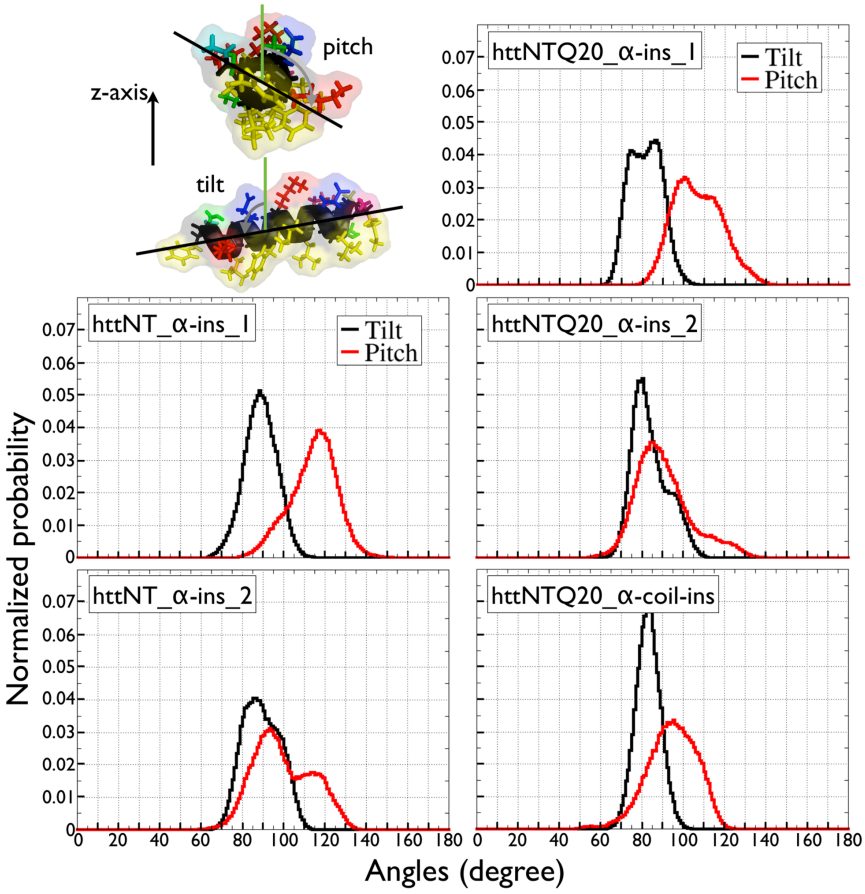


Figure IV.8 – Orientations of the amphipathic plane of the htt^{NT} region for simulations of both htt^{NT} and htt^{NT}Q$_{20}$ starting from an initially inserted peptide. The graphical definition of the tilt and rotational pitch angles is shown on the top left with the z-axis corresponding to the membrane's normal. The other panels depict the ensemble of angles sampled during each simulation on the 200-500 ns time interval.

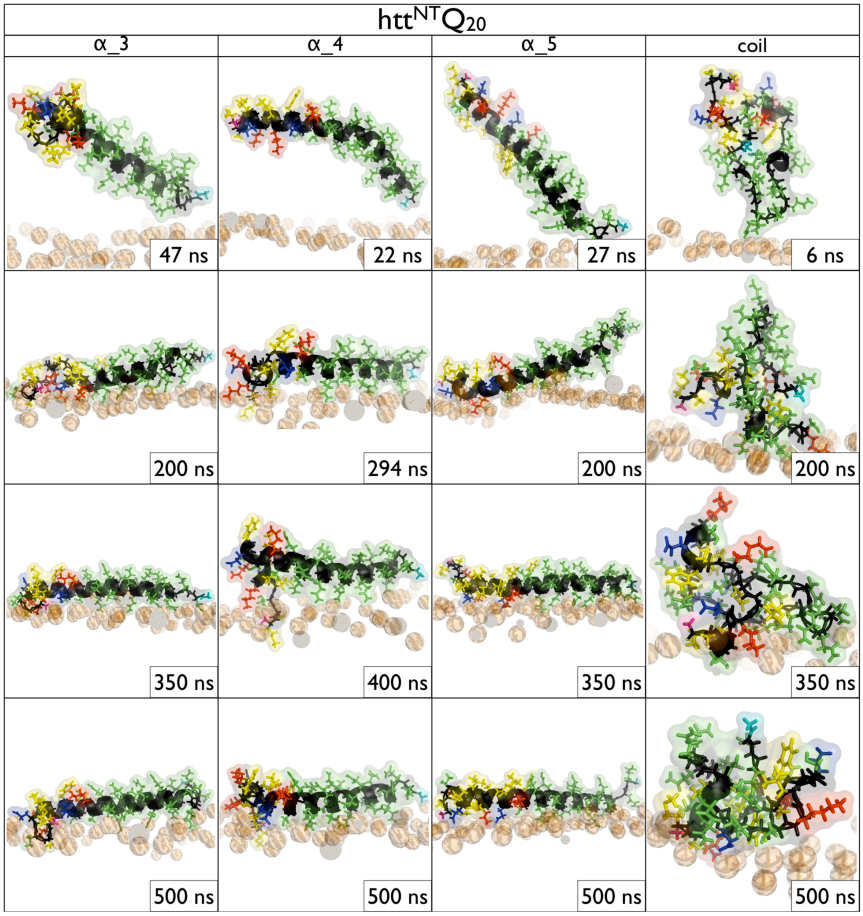


Figure IV.9 – Same as Figure IV.5, for htt^{NT}Q$_{20}$ (simulations httNTQ20_α_3, httNTQ20_α_4, httNTQ20_α_5 and httNTQ20_coil).

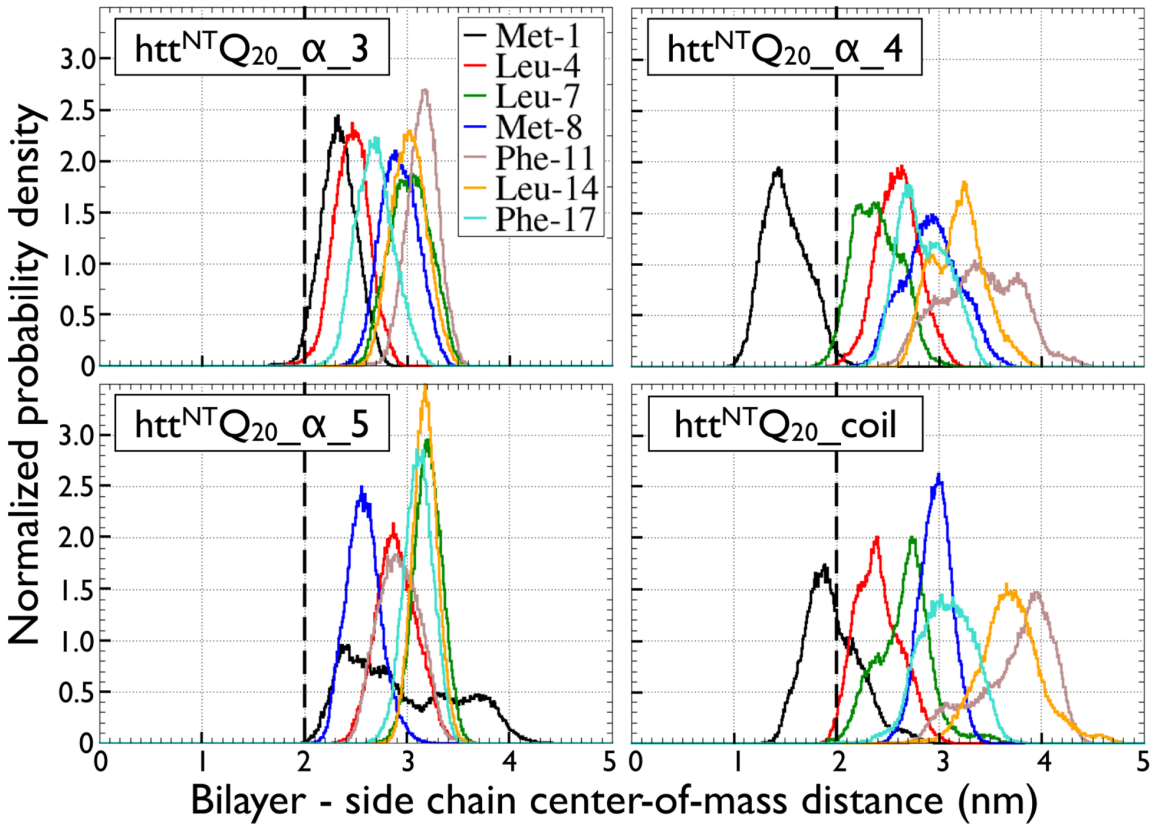


Figure IV.10 – Same as Figure IV.7, for htt^{NT}Q$_{20}$ (simulations httNTQ20_α_3, httNTQ20_α_4, httNTQ20_α_5 and httNTQ20_coil).

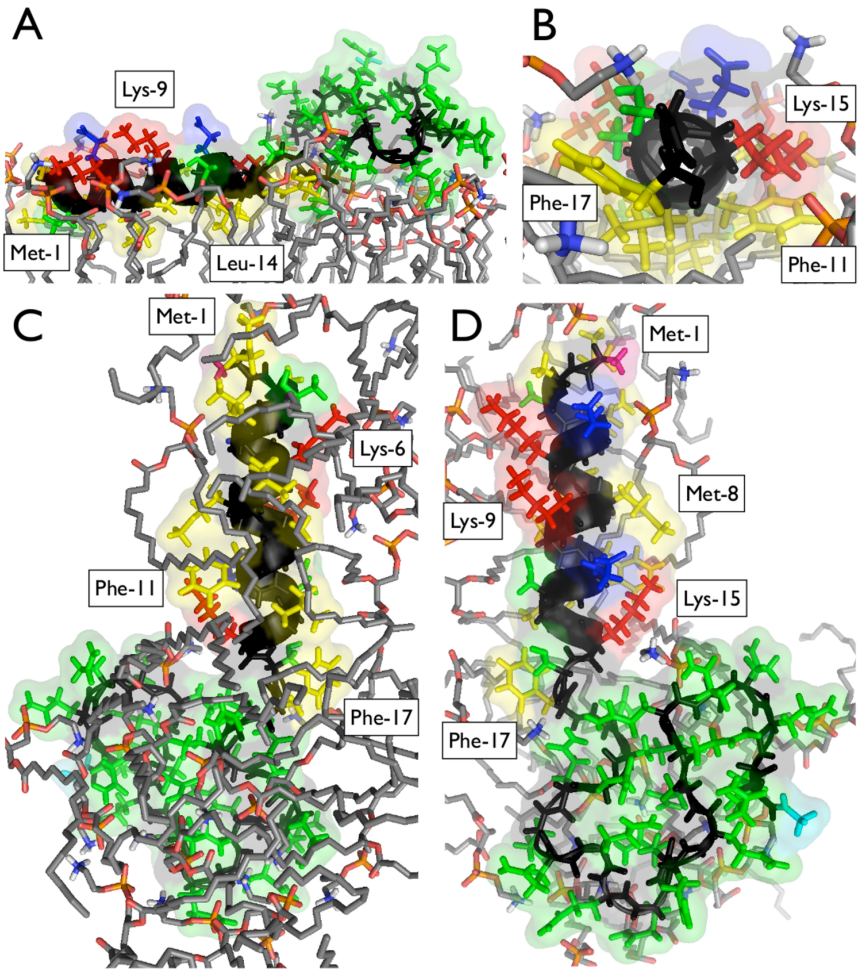


Figure IV.11 – Atomic structure of the average orientation of htt^{NT} over the last 300 ns of httNTQ20_α-coil–ins. (A) Side view, (B) view from the C-terminal , (C) view from the membrane and (D) view from the solvent. For the peptide, its non-polar, negatively charged, positively charged and polar amino acids are respectively shown in yellow, red, blue and green. The backbone atoms are displayed in black, and the N- and C-terminal are respectively shown in pink and teal. For the phospholipids, their phosphorus, oxygen, nitrogen and carbon atoms are respectively displayed in orange, red, blue and grey.

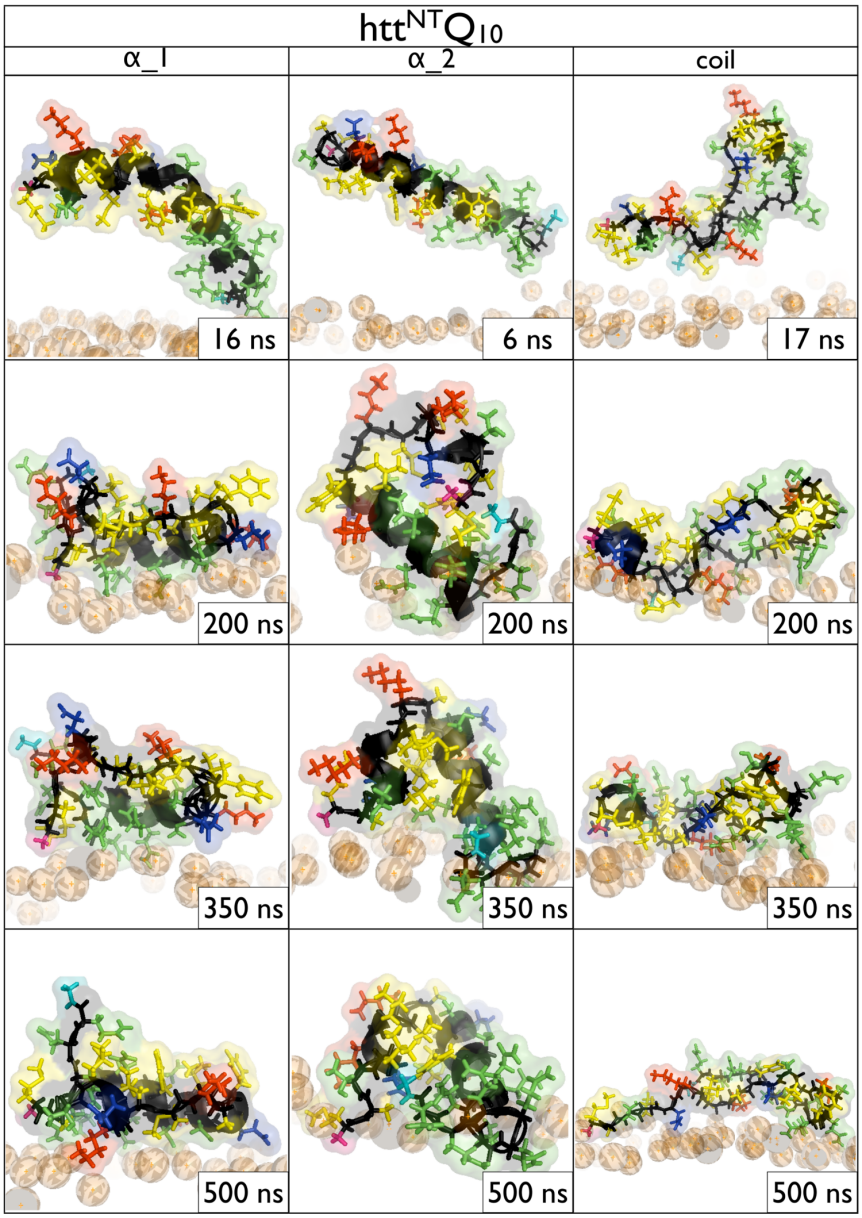


Figure IV.12 – Same as Figure IV.5, for htt^{NT}Q$_{10}$ (simulations httNTQ10_α_1, httNTQ10_α_2 and httNTQ10_coil).

Annexe V

Supporting Information for Chapiter 12

Probing the Huntingtin 1-17 Membrane Anchor on a Phospholipid Bilayer by Using All-atom Simulations

Sébastien Côté†, Vincent Binette†, Evgeniy S. Salnikov‡, Burkhard Bechinger‡, and Normand Mousseau†

† Département de Physique and Groupe d'étude des protéines membranaires (GÉPROM), Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec), Canada

‡ Université de Strasbourg/Center National de la Recherche Scientifique, UMR7177, Institut de Chimie, Strasbourg, France

Materials and Methods

Simulations protocols. All simulations are performed using the Gromacs software (version 4.6.5) [42, 185, 387, 478] combined with the PLUMED plug-in (version 2.0.2) for the HREX simulations [74, 467]. The AMBER99sb*-ILDN protein forcefield [382] is combined with the Berger phospholipid forcefield as previously reported and tested [102]. The following simulation parameters are inspired from previous studies on bilayer systems [102, 384]. The temperature is kept at 303 K using the Bussi–Donadio–Parrinello thermostat with a coupling constant of 0.1 ps [75]. The pressure is controlled semi-isotropically to 1 atm using the Parrinello-Rahman barostat with a coupling constant of 10 ps [348, 368]. Van der Waals and short-range electrostatic interactions are cutoff at 1.0 nm, and long-range electrostatic interactions are computed using smooth Particle-Mesh Ewald (SPME) [116, 150]. Bond lengths and geometry of TIP3P water molecules are respectively constrained using LINCS [184] and SETTLE [326] allowing an integration time step of 2 fs. The center-of-mass motion is removed every 20 fs. The configurations sampled during the MD simulations are saved every 10 ps. For comparison purposes, we also use another phospholipid forcefield, SLIPIDS [203–205, 244], that is compatible with AMBER99sb*-ILDN and for which we use the above cutoff schemes as they are also appro-

priate for it (Joakim P. M. Jämbeck, personal communication).

Hamiltonian replica exchange. To reinforce our observations from MD simulations concerning the configuration of Htt17 on the bilayer, we also use a sampling enhancing algorithm called Hamiltonian replica-exchange (HREX). It belongs to the widely used parallel tempering methods [181, 449]. The REST2 standard of HREX is implemented in PLUMED [74, 492]. This sampling enhancing algorithm relies more specifically on executing simultaneous simulations called replicas with different Hamiltonian (energy) and allowing, at a given frequency, exchanges between replicas i and j respectively at neighboring scales m and n with a probability of

$$P(X_i \leftrightarrow X_j) = \min\left[1, \exp\left(\frac{-H_m(X_j) + H_m(X_i) - H_n(X_i) + H_n(X_j)}{k_B T}\right)\right] \tag{V.1}$$

where H is the energy, X is the position coordinates, T is the temperature and

$$H_m(X) = \lambda_m H_{pp}(X) + (\lambda_m)^{1/2} H_{ps} + H_{ss}(X) \tag{V.2}$$

where H_m is the energy at scale m, H_{pp} is the protein–protein energy, H_{ps} is the protein–solvent energy, H_{ss} is the solvent–solvent energy, and λ_m is the scaling at scale m ($\lambda_m \leq 1.0$). Here, the solvent term consists of the water molecules and the phospholipids. REST2 has been tested on trpcage and a β-hairpin showing a significantly lower computational cost and better sampling than temperature replica exchange [492]. Moreover, in contrast to temperature replica exchange, this protocol allows us to enhance the sampling using only room temperature, which is necessary to conserve the integrity and the physical properties of the phospholipid bilayer.

In our HREX simulation, we use 16 scales generated by a geometric distribution : 1.0, 0.92, 0.85, 0.79, 0.73, 0.67, 0.62, 0.57, 0.53, 0.49, 0.45, 0.41, 0.38, 0.35, 0.33 and 0.3. An exchange between each neighboring scale is attempted every 2 ps resulting in an exchange rate of 10–30%.

Simulated systems. For our simulations, we use two initial states : the NMR model obtained in the presence of DPC micelles (PDB : 2LD2) [319], and a single α-helix. The NMR model that we use is the third structure in the 2LD2 PDB as it best satisfies the chemical shift constraints obtained by solid-state NMR on a POPC bilayer [318]. The other structures are very similar to the third one as they all populate disordered configurations from residues 1 to 4 and an α-helix from residues 6 to 16 [319] according to STRIDE [164]. For its part, the single α-helix, as its name indicates, has a single α-helix running from residues 2 to 16 and is generated from the

equilibrated NMR model – bilayer system. To extend the α-helix, we apply steered molecular dynamics on the α-helical content using PLUMED. While doing so, the positions of the main chain atoms of residues 10 to 17 are restrained to their initial position in order to preserve the same orientation and insertion depth.

Simulations are done in a hexagonal prism periodic cell and the membrane is placed such that its cross-section is hexagonal with its normal pointing towards the z axis. The bilayer contains 84 POPC per leaflet and the minimal distance between the peptide and its periodic images is always greater than 3.0 nm. The center-of-mass of the backbone atoms of residues 6 to 17 is placed at 1.8 nm from the bilayer center-of-mass, and the non-polar residues are facing the hydrophobic core of the membrane. We also tested two others initial positions at 2.0 and 2.5 nm using HREX simulations only. Clashes are first removed using the InflateGRO2 procedure [413] resulting in the removal of 1 POPC on the leaflet where the peptide is. The system is then solvated and neutralized by the addition of two chloride ions, energy minimized using the conjugate-gradient algorithm, and equilibrated in the NPT ensemble at 303K for 100 ns restraining the backbone atoms of the protein to their initial position using a harmonic potential.

Forcefields. The fundamental motivation in our choice of forcefields is that the protein forcefield must correctly samples the dynamics of the Htt17 membrane anchor as well as various other proteins in aqueous solution. Previously, we observed that AMBER99sb*-ILDN [382] reproduces most closely the configurational ensemble of Htt17 as measured experimentally – 10 to 55% α-helix by CD spectroscopy [16, 319, 459, 504] and no stable α-helix by solution NMR [459] – while other popular forcefields either overestimate – AMBER03 [141] and CHARMM27 [299, 300] – or underestimate – OPLS-AA\L [217, 220] – the stability of α-helices for this membrane anchor [111]. Furthermore, it has been demonstrated using microseconds MD simulations that this forcefield is one of the best to reproduce the stability and folding of various ordered and disordered proteins [28, 98, 279, 380, 383]. AMBER99sb*-ILDN originates from slight corrections to the bonded parameters of AMBER99sb [190] on the ψ dihedral angle correcting a helix/coil stability imbalance (AMBER99sb*) [46], and on the χ_1 and χ_2 dihedral angles to correct their energy landscape using quantum mechanical computations (AMBER99sb-ILDN) [281].

Once we determined that AMBER99sb*-ILDN is best suited for our peptide, we chose the Berger phospholipid forcefield because it is compatible with the AMBER family of force-

fields, usable in the NPT ensemble, and nicely reproduce the properties of POPC bilayers. The combination of the Berger and AMBER provides reliable free energy measurements of amino acid transfer from water to cyclohexane as well as good structural and dynamical properties of membrane–protein systems [102]. To confirm our choice of phospholipid forcefield, we also perform other simulations with the SLIPIDS phospholipid forcefield to compare against our main simulations as reported in the Supporting Material. The combination of the AMBER family of forcefields with SLIPIDS reproduces the free energy partitioning of amino acids in cyclohexane and of small molecules in phospholipid bilayers, as well as the tilt angle of the WALP23 membrane spanning peptide in phospholipid bilayers [203–205, 361]. The conclusions related to membrane-protein systems were obtained for proteins parametrized using AMBER99sb or its variants with small corrections to the bonded parameters. As such, these conclusions can be extended to AMBER99sb*-ILDN as it only differs from AMBER99sb by slight modifications to the bonded parameters (ψ, χ_1 and χ_2 dihedral angles), thus not affecting protein–membrane interactions.

Analysis. In-house and Gromacs utilities are used unless otherwise specified. Secondary structures are computed using STRIDE [164]. Hydrogen bonds are considered formed when the donor–acceptor distance is within 3.5 Å and the hydrogen–donor–acceptor angle is smaller than 30°. Salt-bridges are considered formed when the distance between two oppositely charged moieties is within 4 Å [21]. The orientation of Htt17 in the membrane is analyzed using ^{15}N chemical shifts and ^{2}H quadrupolar splitting as these quantities depend on the orientation of the peptide on the bilayer [34, 36, 408]. More precisely, the ^{15}N chemical shift is directly related to the orientation of its tensor eigenvectors with respect to the external magnetic field that is aligned to the membrane normal. The eigenvectors are oriented in a very specific way within the peptide molecular frame. One of them, $\vec{\sigma}_{22}$, is oriented perpendicular to the peptide plane while the other two, $\vec{\sigma}_{11}$ and $\vec{\sigma}_{33}$, are in the peptide plane. The orientation of $\vec{\sigma}_{11}$ is between the N–C and N–H_N bonds at about 16° from the latter bond, and the orientation of $\vec{\sigma}_{33}$ is perpendicular to $\vec{\sigma}_{11}$ [408]. Defining the angle Φ, Θ and Ψ as the Euler angles needed to pass from the eigenvectors system to the laboratory frame, the chemical shift can be then calculated [36] from

$$\sigma_{zz} = \sigma_{11}\sin^2\theta\cos^2\Phi + \sigma_{22}\sin^2\theta\sin^2\Phi + \sigma_{33}\cos^2\theta \tag{V.3}$$

where σ_{zz} is the ^{15}N chemical shift, Φ is the rotation around the z-axis, Θ is the rotation around

the y'-axis, Ψ is the rotation around the z''-axis, and σ_{11}, σ_{22} and σ_{33} are the eigenvalues of the chemical shift tensor that are respectively taken to be 56 ppm, 81 ppm and 223 ppm as done in the NMR experiment on Htt17 [318]. In a similar manner, the quadrupolar splitting is also related to the peptide orientation with respect to the magnetic field. Indeed, the ^{2}H quadrupolar splitting can be calculated [5] from

$$\Delta\nu_Q = \frac{3}{2}\frac{e^2qQ}{h}\frac{3\cos^2\theta - 1}{2} \tag{V.4}$$

where θ is the angle between the C_α–C_β bond and the external magnetic field that is aligned to the membrane normal, and $\frac{e^2qQ}{h}$ is the static quadrupolar coupling constant that is taken to be 74 kHz as done in the NMR experiment on Htt17 [318]. Finally, the structural properties of the membrane such as the area per lipid and the membrane thickness are respectively computed using the VTMC program [329] and the GridMAT script [7], and the order parameters of the acyl chain carbons are computed using

$$S = \langle\frac{3\cos^2(\theta) - 1}{2}\rangle \tag{V.5}$$

where θ is the angle between the C–H bonds vector and the bilayer normal. The average is performed over all phospholipids in all time frames.

Supporting Figures

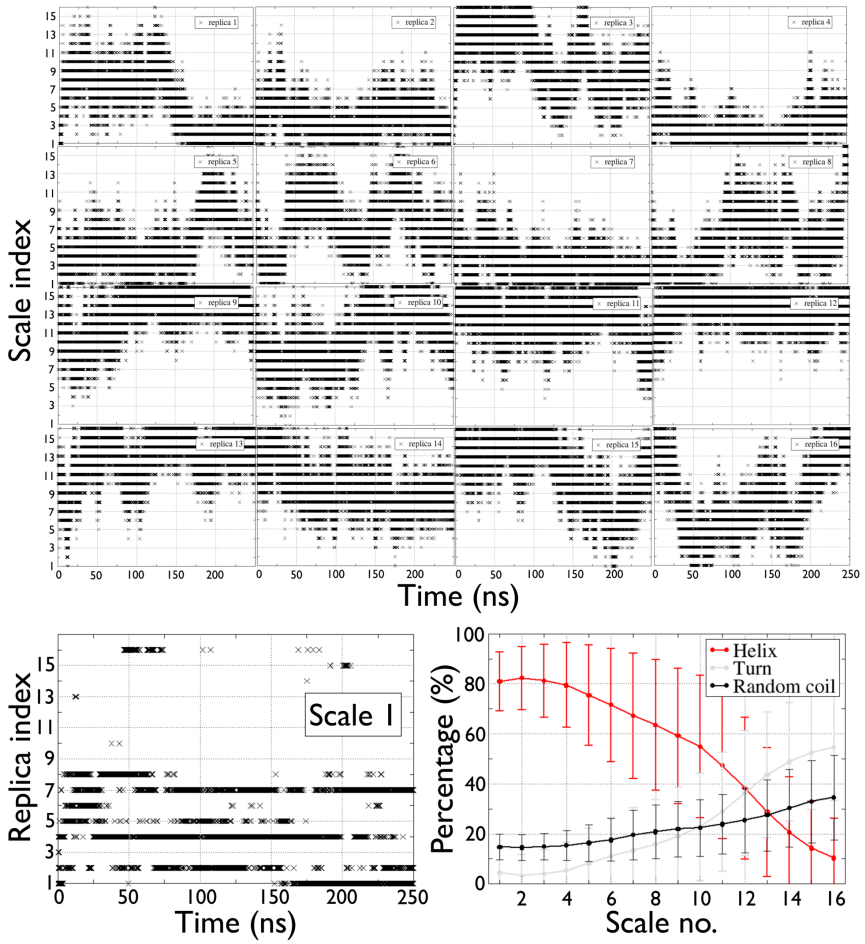


Figure V.1 – Assessment of the sampling of the first HREX simulation starting from a single α-helix (Htt17_α_hrex). Top panel : diffusion from scale to scale for each of the 16 replicas as a function of time. Scale one corresponds to the standard forcefield parameters (unscaled) and scale 16 to the most reduced scale (0.3). Bottom left : replicas visiting the first scale (unscaled energy) as a function of time. Bottom right : secondary structure as a function of the scale number. For the latter graph, the average and standard deviation are computed on the 50–250 ns time interval using 20-ns time windows.

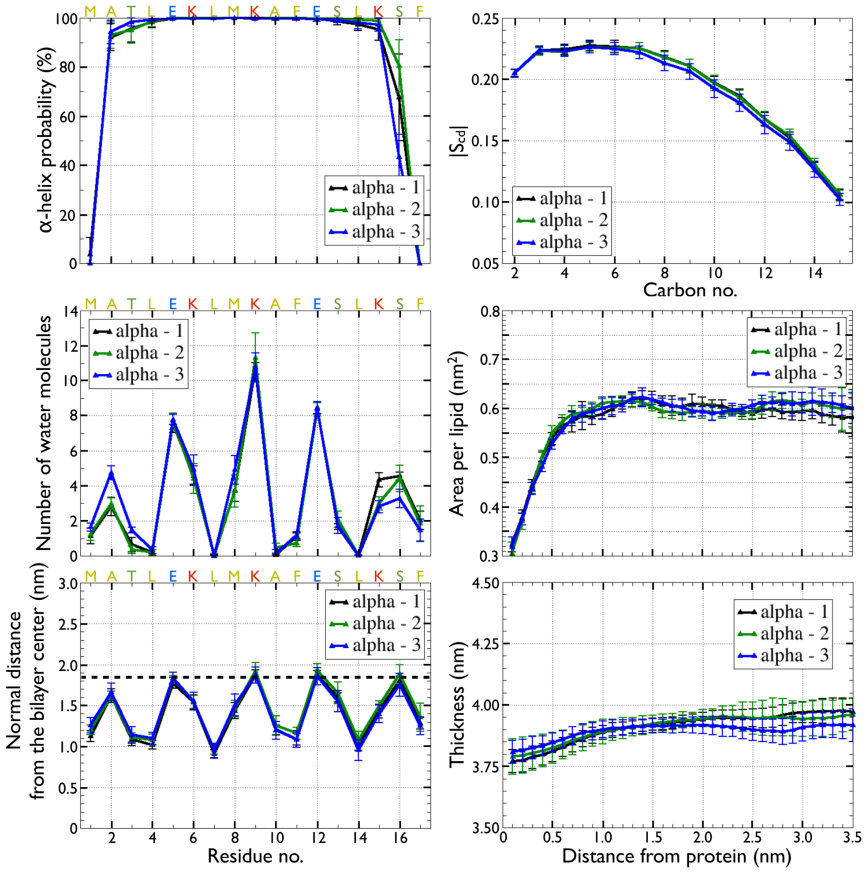


Figure V.2 – Comparison between the HREX simulations starting from a single α-helix at different insertion depths : 1.8 nm (alpha – 1), 2.0 nm (alpha – 2) and 2.5 nm (alpha – 3) from the bilayer's center. To the left, from top to bottom, the per residue α-helix probability, water accessibility and insertion depth are displayed. To the right, from top to bottom, the order parameter of each C-H bonds, as well as the area per lipid and membrane thickness as a function of minimal distance from the protein. The average and standard deviation are computed on the 50–250 ns time interval using 20-ns time windows.

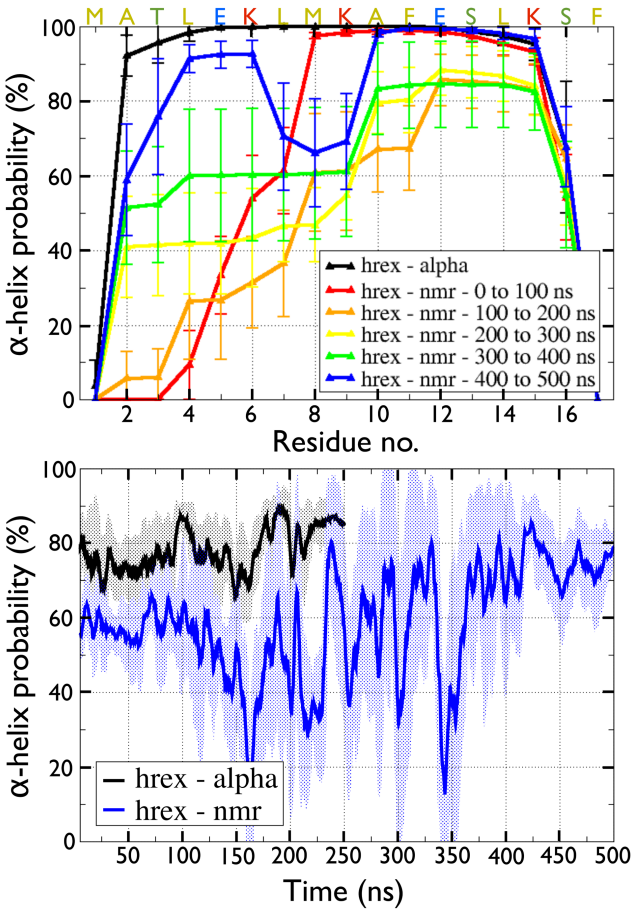


Figure V.3 – Alpha-helix probability at the unscaled replica for the HREX simulation starting from the NMR model (Htt17_nmr_hrex). The per residue average and the global average as a function of time are shown to the top and bottom respectively. For the former graph, the average and standard deviation are computed using 20-ns time windows on the 50–250 ns time interval for the Htt17_α_hrex simulation and on 5 intervals of 100 ns for the Htt17_nmr_hrex simulation. For the latter graph, a running time average (solid line) and standard deviation (dotted region) are displayed using a 5-ns window.

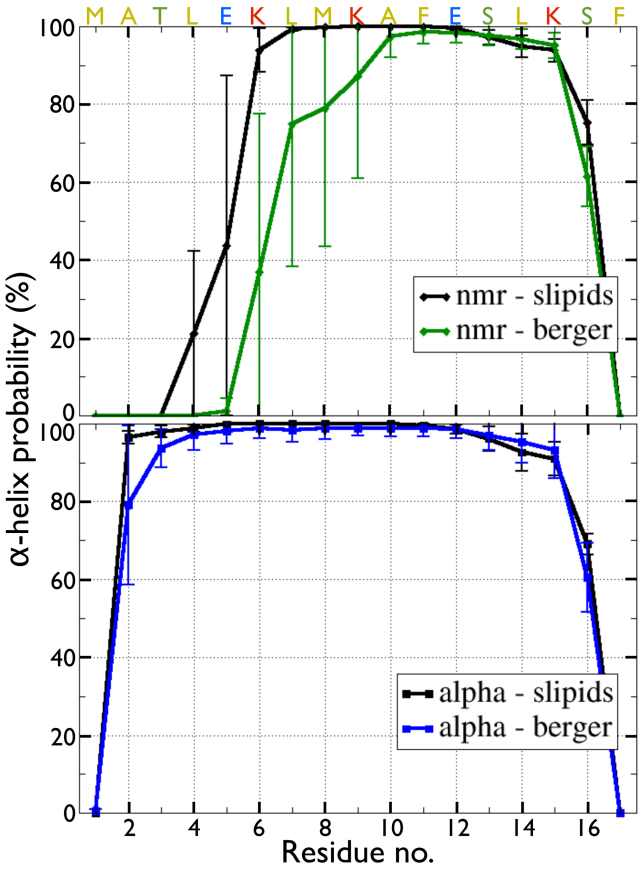


Figure V.4 – Per residue α-helix probability for Htt17. The results for simulations starting from the NMR model and a single α-helix are respectively shown in the top and bottom panels. For each initial state, we compare the simulations using AMBER99sb*-ILDN/Berger (green and blue curves) to AMBER99sb*-ILDN/SLIPIDS (black curves). Initially, according to STRIDE [164], the NMR model has an α-helix from residues 6 to 16 and the alpha model from residues 2 to 16. The value and the error bar for each residue are respectively the average and the standard deviation over all the averages on the 250–1000 ns time interval obtained from 11 (berger) or 2 (slipids) independent MD simulations.

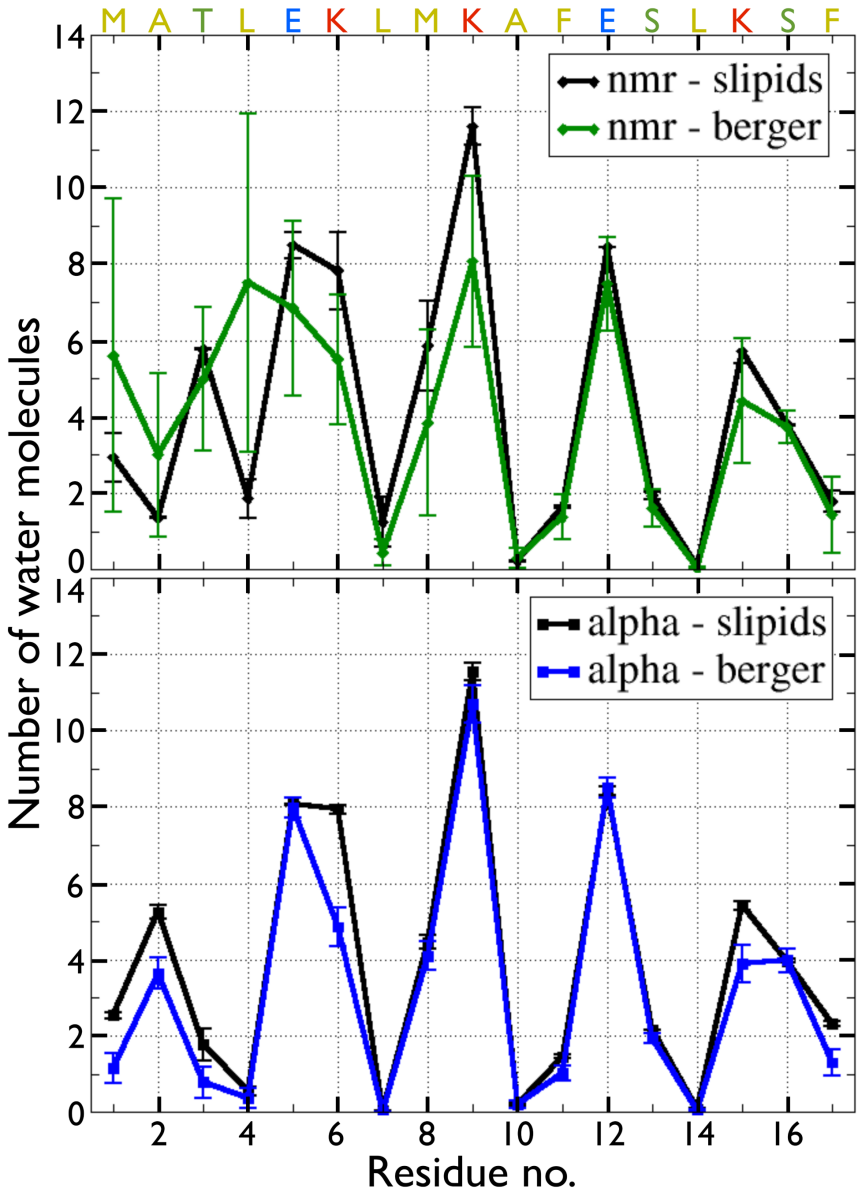


Figure V.5 – Per residue solvent accessibility for Htt17. Results for simulations starting from the NMR model and a single α-helix are respectively shown in the top and bottom panels. For each initial state, we compare the simulations using AMBER99sb*-ILDN/Berger (green and blue curves) to AMBER99sb*-ILDN/SLIPIDS (black curves). The value and the error bar for each residue are respectively the average and the standard deviation over all the averages on the 250–1000 ns time interval obtained from 11 (berger) or 2 (slipids) independent MD simulations.

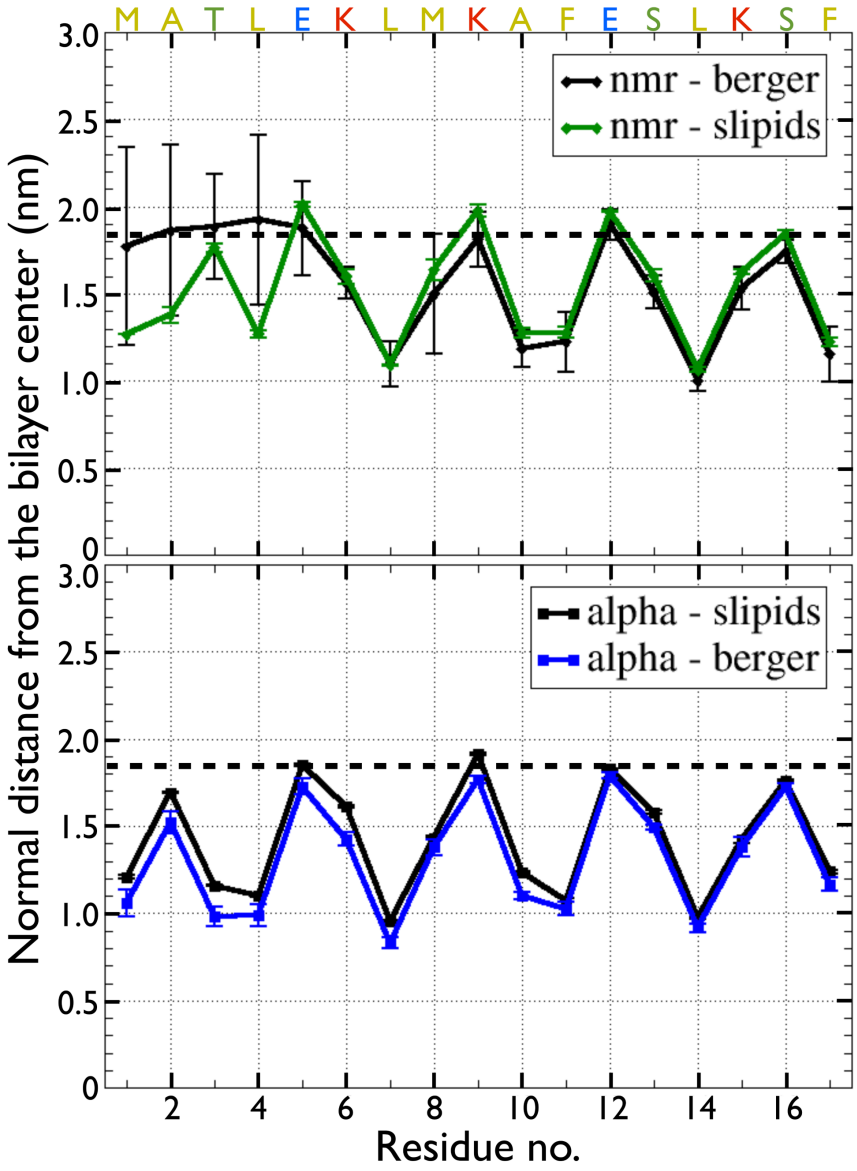


Figure V.6 – Center-of-mass distance between each side chain and the bilayer along the direction perpendicular to the membrane surface (z-axis). Results for simulations starting from the NMR model and a single α-helix are respectively shown in the top and bottom panels. For each initial state, we compare the simulations using AMBER99sb*-ILDN/Berger (green and blue curves) to AMBER99sb*-ILDN/SLIPIDS (black curves). The dotted line at 1.8 nm is the average position of the phosphorous atoms of the upper leaflet with respect to the center of the bilayer. The value and the error bar for each residue are respectively the average and the standard deviation over all the averages on the 250–1000 ns time interval obtained from 11 (berger) or 2 (slipids) independent MD simulations.

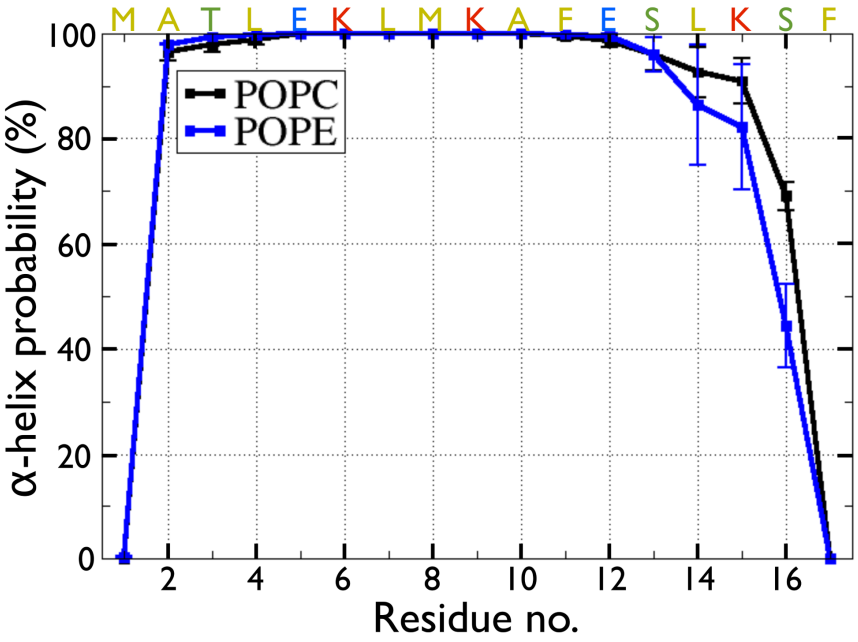


Figure V.7 – Comparison between POPC (black) and POPE (blue) bilayers in terms of the per residue α-helix probability of Htt17. These simulations were done using the AMBER99*-ILDN/SLIPIDS force-field. The value and the error bar for each residue are respectively the average and the standard deviation over all the averages on the 250–1000 ns (POPC) and 250–500 ns (POPE) time interval obtained from two MD independent simulations.

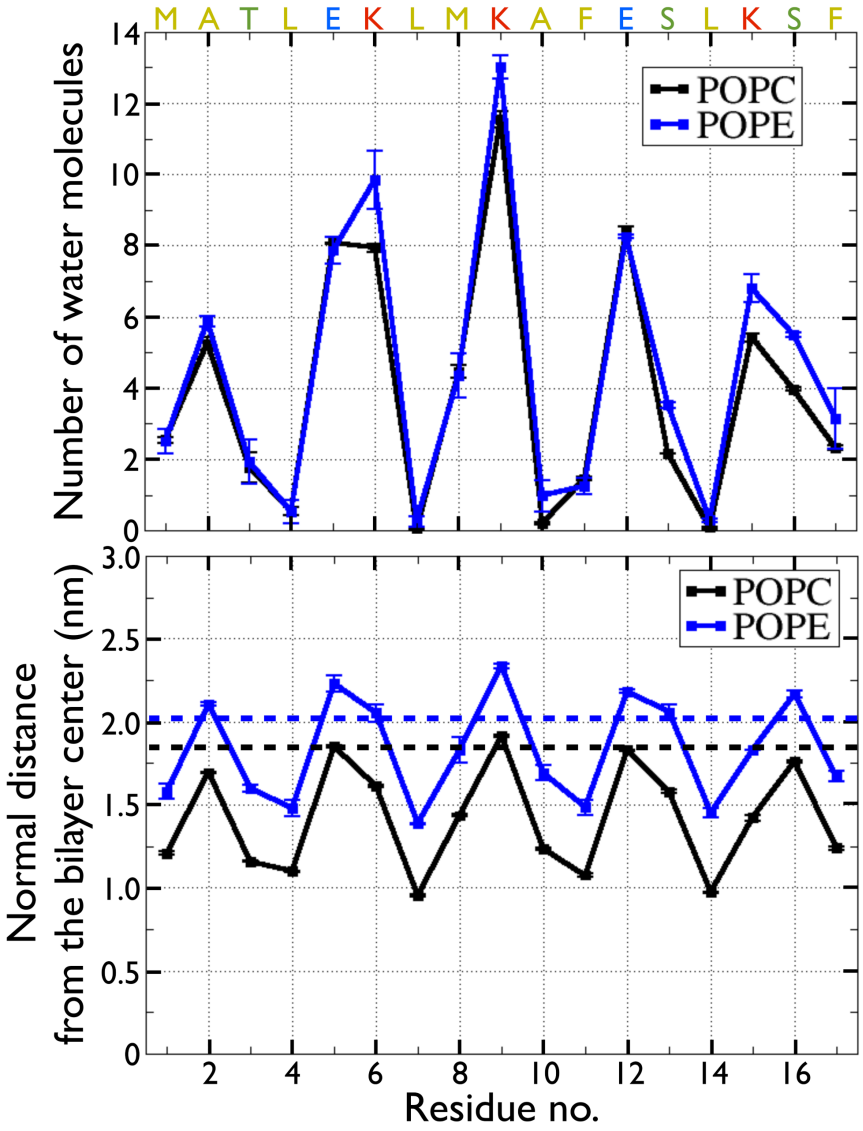


Figure V.8 – Comparison between POPC (black) and POPE (blue) bilayers in terms of solvent accessibility and insertion depth of Htt17. Top panel : The solvent accessibility is quantified in terms of the number of water molecules within 0.35 nm from each residue. Bottom panel : The insertion depth corresponds to the center-of-mass distance between each side chain and the bilayer along the direction perpendicular to the membrane surface (z-axis). The dotted lines are the average position of the phosphorous atoms of the upper leaflet with respect to the center of the POPC (black) and POPE (blue) bilayers. These simulations were done using the AMBER99*-ILDN/SLIPIDS forcefield. The value and the error bar for each residue are respectively the average and the standard deviation over all the averages on the 250–1000 ns (POPC) and 250–500 ns (POPE) time interval obtained from two independent MD simulations.

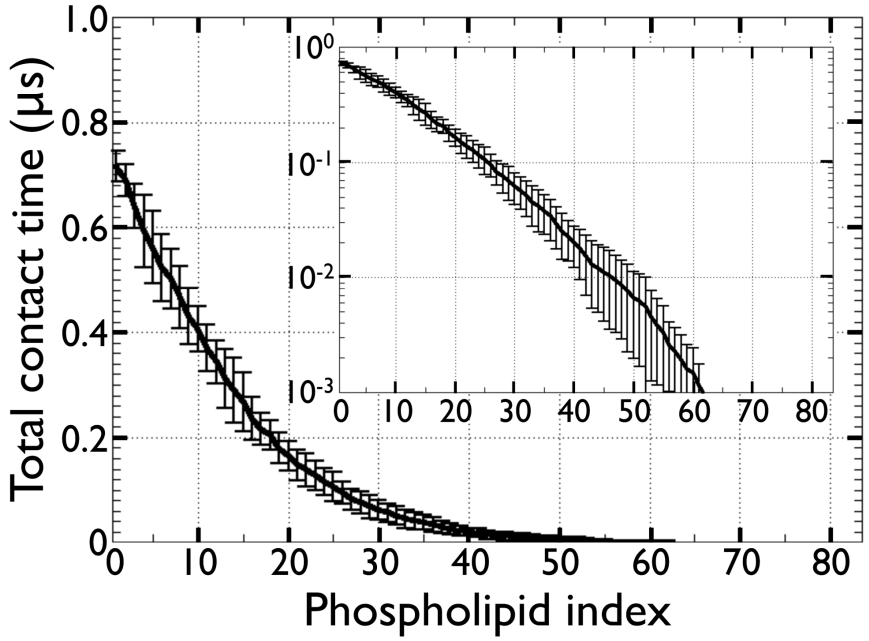


Figure V.9 – Total contact time between each phospholipid and Htt17. The inset depicts the same data using a log scale. The value and the error bar are respectively the average and the standard deviation over all the averages on the 250–1000 ns time interval obtained from 11 independent Htt17_α simulations.

Supporting Tables

Table V.I – Orientation of Htt17 on the bilayer.

	$^{2}H_3$-Ala^{2} kHz	^{15}N-Leu^{7} ppm	$^{2}H_3$-Ala^{10} kHz	^{15}N-Phe^{11} ppm	^{15}N-Leu^{14} ppm	^{15}N-Phe^{17} ppm
Htt17_nmr	14 ± 20	110 ± 28	26 ± 12	91 ± 3	77 ± 7	80 ± 15
Run 1	-22 ± 16	105 ± 26	18 ± 19	95 ± 18	84 ± 24	74 ± 21
Run 2	30 ± 38	90 ± 19	28 ± 20	94 ± 16	72 ± 12	76 ± 19
Run 3	45 ± 17	143 ± 34	58 ± 14	89 ± 12	76 ± 13	87 ± 29
Run 4	14 ± 22	83 ± 14	24 ± 17	91 ± 13	75 ± 13	70 ± 14
Run 5	-1 ± 29	153 ± 31	16 ± 17	97 ± 17	80 ± 19	75 ± 23
Run 6	-1 ± 26	78 ± 17	19 ± 19	89 ± 13	66 ± 10	84 ± 20
Run 7	-5 ± 36	97 ± 20	17 ± 18	94 ± 16	76 ± 13	69 ± 15
Run 8	45 ± 21	105 ± 33	36 ± 21	86 ± 9	68 ± 9	97 ± 27
Run 9	13 ± 28	160 ± 35	25 ± 19	89 ± 17	90 ± 19	67 ± 15
Run 10	14 ± 36	108 ± 25	19 ± 18	88 ± 16	90 ± 18	66 ± 12
Run 11	25 ± 34	81 ± 17	29 ± 19	90 ± 14	74 ± 14	118 ± 36
Htt17_nmr_slipids	22 ± 6	83 ± 2	28 ± 2	90 ± 1	73 ± 2	72 ± 2
Run 1	28 ± 33	80 ± 16	26 ± 19	89 ± 12	71 ± 11	75 ± 17
Run 2	17 ± 36	85 ± 15	30 ± 20	91 ± 14	75 ± 14	70 ± 14
Htt17_α	-7 ± 13	71 ± 2	30 ± 6	85 ± 1	66 ± 1	87 ± 4
Run 1	-1 ± 22	69 ± 8	29 ± 18	85 ± 8	65 ± 7	86 ± 22
Run 2	-14 ± 20	69 ± 8	22 ± 18	85 ± 9	64 ± 6	90 ± 22
Run 3	-1 ± 21	71 ± 10	22 ± 19	87 ± 10	66 ± 8	84 ± 17
Run 4	-19 ± 22	70 ± 8	35 ± 19	84 ± 7	65 ± 7	88 ± 21
Run 5	29 ± 26	77 ± 12	36 ± 15	87 ± 9	68 ± 8	82 ± 21
Run 6	-14 ± 22	71 ± 10	28 ± 17	86 ± 9	66 ± 8	88 ± 23
Run 7	-17 ± 19	71 ± 9	25 ± 18	86 ± 9	64 ± 7	86 ± 20
Run 8	-15 ± 20	70 ± 9	29 ± 20	85 ± 8	66 ± 8	92 ± 26
Run 9	-18 ± 18	71 ± 8	39 ± 19	85 ± 8	68 ± 9	92 ± 21
Run 10	-8 ± 20	72 ± 12	29 ± 22	84 ± 8	66 ± 8	86 ± 24
Run 11	-1 ± 21	73 ± 9	37 ± 15	86 ± 8	67 ± 7	81 ± 17
Htt17_α_slipids	-9 ± 3	71 ± 1	17 ± 1	88 ± 1	66 ± 1	86 ± 2
Run 1	-11 ± 19	71 ± 11	17 ± 19	87 ± 11	66 ± 10	84 ± 22
Run 2	-6 ± 23	71 ± 11	17 ± 20	88 ± 12	67 ± 10	88 ± 26
Htt17_α_hrex	-4 ± 7	68 ± 3	27 ± 8	84 ± 1	65 ± 2	92 ± 5
Htt17_α_hrex_2	-17 ± 10	70 ± 5	28 ± 9	84 ± 2	65 ± 2	92 ± 14
Htt17_α_hrex_3	-15 ± 9	72 ± 6	23 ± 10	87 ± 3	67 ± 4	85 ± 9
ssNMR (exp)	5 ± 2	71.2 ± 1.7	11 ± 2.5	78.9 ± 1.5	73.3 ± 1.2	88.2 ± 0.9

The orientation of the membrane anchor is directly related to the ^{15}N chemical shifts of Leu7, Phe11, Leu14 and Phe17 as well as the $^{2}H_3$ quadrupolar splitting of Ala2 and Ala10 [5, 36, 408]. The last row shows the solid-state NMR measurements of Htt17 on a POPC bilayer [318, 319]. These quantities are computed over the 250–1000 ns time interval for the MD simulations. The first line for each simulation type combines the averages of all independent MD runs. For the HREX simulations, the average and standard deviation are computed at the unscaled replica on the 50–250 ns time interval using 20-ns time windows.

Table V.II – Average number of hydrogen bonds between Htt17 and the phospholipids of the membrane.

	Thr3 no. (%)	Lys6 no. (%)	Lys9 no. (%)	Ser13 no. (%)	Lys15 no. (%)	Ser16 no. (%)
Htt17_nmr	0.9 ± 0.3 (14 ± 19)	1.8 ± 0.3 (85 ± 17)	1.4 ± 0.2 (56 ± 19)	1.0 ± 0.0 (45 ± 22)	1.8 ± 0.3 (86 ± 8)	1.0 ± 0.0 (9 ± 5)
Run 1	0.0 ± 0.0 (0)	1.9 ± 0.7 (94)	1.4 ± 0.6 (72)	1.0 ± 0.1 (13)	1.9 ± 0.8 (82)	1.0 ± 0.1 (4)
Run 2	1.0 ± 0.1 (4)	1.8 ± 0.7 (91)	1.3 ± 0.6 (47)	1.0 ± 0.1 (65)	2.0 ± 0.7 (92)	1.0 ± 0.1 (3)
Run 3	1.0 ± 0.1 (11)	1.0 ± 0.2 (41)	1.5 ± 0.7 (43)	1.0 ± 0.1 (95)	2.0 ± 0.7 (91)	1.0 ± 0.1 (18)
Run 4	1.0 ± 0.1 (4)	1.7 ± 0.7 (92)	1.6 ± 0.7 (52)	1.0 ± 0.1 (37)	1.6 ± 0.6 (91)	1.0 ± 0.1 (5)
Run 5	1.0 ± 0.1 (67)	1.7 ± 0.6 (89)	1.3 ± 0.5 (52)	1.0 ± 0.1 (19)	1.7 ± 0.6 (85)	1.0 ± 0.1 (10)
Run 6	1.0 ± 0.1 (9)	2.0 ± 0.8 (91)	1.2 ± 0.4 (18)	1.0 ± 0.1 (50)	1.7 ± 0.7 (89)	1.0 ± 0.1 (6)
Run 7	0.0 ± 0.0 (0)	2.0 ± 0.8 (95)	1.6 ± 0.8 (64)	1.0 ± 0.1 (49)	1.6 ± 0.6 (84)	1.0 ± 0.1 (7)
Run 8	1.0 ± 0.1 (26)	1.4 ± 0.6 (61)	1.2 ± 0.5 (68)	1.0 ± 0.1 (59)	2.2 ± 0.8 (88)	1.0 ± 0.1 (4)
Run 9	1.0 ± 0.1 (1)	1.9 ± 0.7 (94)	1.8 ± 0.6 (94)	1.0 ± 0.1 (39)	1.5 ± 0.7 (71)	1.0 ± 0.1 (11)
Run 10	1.0 ± 0.1 (27)	2.1 ± 0.7 (96)	1.6 ± 0.7 (63)	1.0 ± 0.1 (22)	1.5 ± 0.7 (72)	1.0 ± 0.1 (17)
Run 11	1.0 ± 0.1 (10)	2.0 ± 0.8 (94)	1.2 ± 0.4 (40)	1.0 ± 0.1 (48)	2.6 ± 0.6 (99)	1.0 ± 0.1 (13)
Htt17_nmr_slipids	1.0 ± 0.0 (59 ± 1)	1.4 ± 0.1 (71 ± 6)	1.1 ± 0.0 (10 ± 3)	1.0 ± 0.0 (48 ± 8)	1.3 ± 0.0 (66 ± 3)	1.0 ± 0.0 (12 ± 1)
Run 1	1.0 ± 0.1 (60)	1.5 ± 0.7 (77)	1.1 ± 0.2 (8)	1.0 ± 0.1 (56)	1.3 ± 0.5 (69)	1.0 ± 0.1 (11)
Run 2	1.0 ± 0.1 (57)	1.4 ± 0.5 (65)	1.1 ± 0.2 (13)	1.0 ± 0.1 (40)	1.3 ± 0.5 (63)	1.0 ± 0.1 (13)
Htt17_α	1.0 ± 0.0 (61 ± 30)	1.9 ± 0.2 (92 ± 4)	1.4 ± 0.2 (38 ± 11)	1.0 ± 0.0 (54 ± 8)	1.7 ± 0.1 (90 ± 4)	1.0 ± 0.0 (6 ± 2)
Run 1	1.0 ± 0.1 (85)	1.7 ± 0.7 (90)	1.3 ± 0.5 (24)	1.0 ± 0.1 (60)	1.6 ± 0.6 (86)	1.0 ± 0.1 (5)
Run 2	1.0 ± 0.1 (29)	2.1 ± 0.8 (94)	1.2 ± 0.4 (27)	1.0 ± 0.1 (59)	1.8 ± 0.5 (94)	1.0 ± 0.1 (4)
Run 3	1.0 ± 0.1 (92)	1.7 ± 0.7 (91)	1.4 ± 0.7 (39)	1.0 ± 0.1 (51)	1.8 ± 0.5 (92)	1.0 ± 0.1 (7)
Run 4	1.0 ± 0.1 (14)	2.1 ± 0.8 (92)	1.2 ± 0.4 (25)	1.0 ± 0.1 (51)	1.7 ± 0.8 (87)	1.0 ± 0.1 (8)
Run 5	1.0 ± 0.1 (73)	2.3 ± 0.7 (98)	1.3 ± 0.6 (45)	1.0 ± 0.1 (60)	1.6 ± 0.6 (86)	1.0 ± 0.1 (3)
Run 6	1.0 ± 0.1 (65)	1.9 ± 0.8 (92)	1.3 ± 0.6 (34)	1.0 ± 0.1 (50)	1.7 ± 0.6 (93)	1.0 ± 0.1 (7)
Run 7	1.0 ± 0.1 (87)	1.7 ± 0.7 (82)	1.9 ± 0.8 (42)	1.0 ± 0.1 (67)	1.7 ± 0.6 (91)	1.0 ± 0.1 (6)
Run 8	1.0 ± 0.1 (39)	1.8 ± 0.8 (89)	1.3 ± 0.5 (40)	1.0 ± 0.1 (35)	1.8 ± 0.7 (88)	1.0 ± 0.1 (8)
Run 9	1.0 ± 0.1 (13)	1.5 ± 0.6 (89)	1.2 ± 0.4 (41)	1.0 ± 0.1 (45)	1.8 ± 0.8 (85)	1.0 ± 0.1 (4)
Run 10	1.0 ± 0.1 (86)	2.2 ± 0.8 (94)	1.4 ± 0.7 (39)	1.0 ± 0.1 (54)	1.7 ± 0.7 (91)	1.0 ± 0.1 (10)
Run 11	1.0 ± 0.1 (90)	2.0 ± 0.7 (98)	1.7 ± 0.7 (67)	1.0 ± 0.1 (58)	1.8 ± 0.6 (92)	1.0 ± 0.1 (5)
Htt17_α_slipids	1.0 ± 0.0 (37 ± 21)	1.3 ± 0.0 (63 ± 4)	1.0 ± 0.0 (9 ± 1)	1.0 ± 0.0 (47 ± 10)	1.3 ± 0.0 (63 ± 1)	1.0 ± 0.0 (10 ± 1)
Run 1	1.0 ± 0.1 (58)	1.3 ± 0.5 (59)	1.0 ± 0.2 (10)	1.0 ± 0.1 (38)	1.3 ± 0.5 (63)	1.0 ± 0.1 (10)
Run 2	1.0 ± 0.1 (17)	1.4 ± 0.6 (67)	1.0 ± 0.1 (9)	1.0 ± 0.1 (57)	1.2 ± 0.5 (64)	1.0 ± 0.1 (9)
Htt17_α_hrex	1.0 ± 0.0 (51 ± 30)	2.2 ± 0.2 (95 ± 5)	1.1 ± 0.1 (32 ± 9)	1.0 ± 0.1 (54 ± 19)	1.7 ± 0.2 (90 ± 3)	1.0 ± 0.1 (5 ± 4)
Htt17_α_hrex_2	1.0 ± 0.0 (75 ± 12)	2.1 ± 0.2 (97 ± 2)	1.2 ± 0.1 (24 ± 21)	1.0 ± 0.1 (30 ± 20)	2.2 ± 0.2 (98 ± 2)	1.0 ± 0.1 (3 ± 2)
Htt17_α_hrex_3	1.0 ± 0.0 (51 ± 35)	1.7 ± 0.1 (92 ± 5)	1.4 ± 0.2 (42 ± 14)	1.0 ± 0.1 (71 ± 14)	1.8 ± 0.1 (96 ± 3)	1.0 ± 0.1 (11 ± 9)

For a given residue, statistics are performed only over the frames in which it forms at least one H-bond. The numbers in parenthesis represent the probability of occurrence in percentage during each simulation. These quantities are computed over the 250–1000 ns time interval for the MD simulations. The first line for each simulation type combines the averages of all independent MD runs. For the HREX simulations, the average and standard deviation are computed at the unscaled replica on the 50–250 ns time interval using 20-ns time windows.

Table V.III – Average number of salt-bridges between Htt17 and the phospholipids of the membrane.

	Glu5 no. (%)	Glu12 no. (%)	Lys6 no. (%)	Lys9 no. (%)	Lys15 no. (%)
Htt17_nmr	0.9 ± 0.3 (1 ± 2)	0.9 ± 0.3 (1 ± 1)	1.1 ± 0.1 (49 ± 11)	1.2 ± 0.1 (40 ± 13)	1.1 ± 0.1 (51 ± 8)
Run 1	1.0 ± 0.0 (1)	0.0 ± 0.0 (0)	1.3 ± 0.5 (68)	1.2 ± 0.4 (36)	1.1 ± 0.3 (38)
Run 2	1.0 ± 0.0 (1)	1.0 ± 0.0 (1)	1.1 ± 0.3 (52)	1.1 ± 0.2 (34)	1.2 ± 0.4 (64)
Run 3	1.0 ± 0.0 (1)	1.0 ± 0.0 (1)	1.0 ± 0.1 (31)	1.1 ± 0.3 (31)	1.2 ± 0.4 (60)
Run 4	0.0 ± 0.0 (0)	1.0 ± 0.0 (1)	1.1 ± 0.4 (49)	1.1 ± 0.3 (36)	1.1 ± 0.3 (59)
Run 5	1.0 ± 0.0 (1)	1.0 ± 0.0 (1)	1.1 ± 0.3 (36)	1.4 ± 0.5 (47)	1.1 ± 0.3 (45)
Run 6	1.0 ± 0.0 (2)	1.0 ± 0.0 (1)	1.1 ± 0.4 (40)	1.1 ± 0.3 (16)	1.1 ± 0.3 (44)
Run 7	1.0 ± 0.0 (1)	1.0 ± 0.0 (1)	1.1 ± 0.3 (60)	1.1 ± 0.3 (45)	1.1 ± 0.3 (47)
Run 8	1.0 ± 0.0 (1)	1.0 ± 0.0 (1)	1.1 ± 0.3 (45)	1.2 ± 0.4 (56)	1.2 ± 0.4 (59)
Run 9	1.0 ± 0.0 (8)	1.0 ± 0.0 (1)	1.1 ± 0.3 (64)	1.3 ± 0.5 (68)	1.1 ± 0.3 (57)
Run 10	1.0 ± 0.0 (1)	1.0 ± 0.0 (1)	1.1 ± 0.3 (41)	1.1 ± 0.3 (42)	1.1 ± 0.3 (45)
Run 11	1.0 ± 0.0 (1)	1.0 ± 0.0 (1)	1.2 ± 0.4 (55)	1.1 ± 0.3 (31)	1.0 ± 0.1 (46)
Htt17_nmr_slipids	1.0 ± 0.0 (3 ± 1)	1.0 ± 0.0 (2 ± 1)	1.2 ± 0.0 (56 ± 5)	1.0 ± 0.0 (10 ± 2)	1.2 ± 0.0 (57 ± 6)
Run 1	1.0 ± 0.1 (4)	1.0 ± 0.1 (1)	1.2 ± 0.4 (60)	1.0 ± 0.2 (8)	1.2 ± 0.4 (63)
Run 2	1.0 ± 0.1 (3)	1.0 ± 0.0 (2)	1.2 ± 0.4 (51)	1.0 ± 0.2 (12)	1.2 ± 0.4 (52)
Htt17_α	1.0 ± 0.0 (1 ± 1)	1.0 ± 0.0 (1 ± 0)	1.2 ± 0.1 (60 ± 8)	1.1 ± 0.0 (28 ± 5)	1.1 ± 0.1 (48 ± 9)
Run 1	1.0 ± 0.0 (3)	1.0 ± 0.0 (1)	1.1 ± 0.4 (50)	1.1 ± 0.2 (20)	1.1 ± 0.3 (46)
Run 2	1.0 ± 0.0 (1)	1.0 ± 0.0 (1)	1.2 ± 0.4 (64)	1.0 ± 0.2 (21)	1.0 ± 0.2 (57)
Run 3	1.0 ± 0.0 (1)	1.0 ± 0.0 (1)	1.1 ± 0.3 (60)	1.1 ± 0.2 (32)	1.1 ± 0.3 (57)
Run 4	1.0 ± 0.0 (2)	1.0 ± 0.0 (1)	1.2 ± 0.4 (58)	1.0 ± 0.2 (20)	1.2 ± 0.4 (47)
Run 5	1.0 ± 0.0 (1)	1.0 ± 0.0 (1)	1.3 ± 0.5 (75)	1.1 ± 0.3 (31)	1.1 ± 0.2 (48)
Run 6	1.0 ± 0.0 (1)	1.0 ± 0.0 (1)	1.2 ± 0.4 (57)	1.1 ± 0.3 (27)	1.1 ± 0.3 (51)
Run 7	1.0 ± 0.0 (2)	1.0 ± 0.0 (1)	1.2 ± 0.4 (56)	1.1 ± 0.4 (29)	1.0 ± 0.2 (41)
Run 8	1.0 ± 0.0 (1)	1.0 ± 0.0 (1)	1.2 ± 0.4 (65)	1.1 ± 0.3 (28)	1.2 ± 0.4 (57)
Run 9	1.0 ± 0.0 (2)	1.0 ± 0.0 (1)	1.1 ± 0.2 (50)	1.0 ± 0.2 (32)	1.1 ± 0.3 (35)
Run 10	1.0 ± 0.0 (1)	1.0 ± 0.0 (1)	1.3 ± 0.4 (57)	1.1 ± 0.4 (31)	1.1 ± 0.2 (32)
Run 11	1.0 ± 0.0 (1)	1.0 ± 0.0 (1)	1.0 ± 0.2 (72)	1.1 ± 0.3 (36)	1.1 ± 0.3 (61)
Htt17_α_slipids	1.0 ± 0.0 (3 ± 1)	1.0 ± 0.0 (2 ± 1)	1.2 ± 0.0 (51 ± 3)	1.0 ± 0.0 (10 ± 1)	1.1 ± 0.0 (46 ± 1)
Run 1	1.0 ± 0.1 (3)	1.0 ± 0.1 (2)	1.2 ± 0.4 (48)	1.0 ± 0.2 (10)	1.1 ± 0.4 (45)
Run 2	1.0 ± 0.1 (2)	1.0 ± 0.1 (1)	1.2 ± 0.4 (54)	1.0 ± 0.1 (10)	1.1 ± 0.4 (47)
Htt17_α_hrex	1.0 ± 0.1 (1 ± 1)	1.0 ± 0.0 (1 ± 1)	1.3 ± 0.2 (65 ± 20)	1.1 ± 0.1 (28 ± 7)	1.0 ± 0.1 (38 ± 10)
Htt17_α_hrex_2	1.0 ± 0.1 (2 ± 2)	1.0 ± 0.0 (1 ± 1)	1.2 ± 0.2 (63 ± 18)	1.1 ± 0.1 (16 ± 14)	1.2 ± 0.1 (54 ± 14)
Htt17_α_hrex_3	1.0 ± 0.0 (1 ± 1)	1.0 ± 0.0 (1 ± 1)	1.4 ± 0.3 (63 ± 18)	1.2 ± 0.1 (36 ± 14)	1.0 ± 0.1 (58 ± 8)

For a given residue, statistics are performed only over the frames in which it forms at least one salt-bridge. The numbers in parenthesis represent the probability of occurrence in percentage during each simulation. These quantities are computed over the 250–1000 ns time interval for the MD simulations. The first line for each simulation type combines the averages of all MD runs. For the HREX simulations, the average and standard deviation are computed at the unscaled replica on the 50–250 ns time interval using 20-ns time windows.

Table V.IV – Probability of occurrence of intramolecular salt-bridges in Htt17.

	Glu5 / Lys6 %	Glu5 / Lys9 %	Glu5 / Lys15 %	Glu12 / Lys6 %	Glu12 / Lys9 %	Glu12 / Lys15 %
Htt17_nmr	1 ± 2	3 ± 8	0 ± 0	5 ± 12	18 ± 12	14 ± 14
Run 1	0	0	0	0	0	0
Run 2	0	0	0	0	7	15
Run 3	2	1	0	42	26	3
Run 4	0	0	0	0	21	16
Run 5	0	28	0	0	36	11
Run 6	1	0	0	0	22	21
Run 7	0	0	0	0	5	10
Run 8	8	5	0	15	27	4
Run 9	2	0	0	0	0	53
Run 10	1	0	0	0	22	21
Run 11	0	2	0	0	29	0
Htt17_nmr_slipids	3 ± 2	14 ± 5	0 ± 0	0 ± 0	22 ± 3	20 ± 3
Run 1	5	18	0	0	25	23
Run 2	1	9	0	0	20	17
Htt17_α	1 ± 1	8 ± 3	0 ± 0	0 ± 0	29 ± 8	8 ± 3
Run 1	1	9	0	0	40	6
Run 2	2	8	0	0	32	7
Run 3	0	8	0	0	31	7
Run 4	2	14	0	0	27	7
Run 5	1	11	0	0	16	14
Run 6	1	7	0	0	33	5
Run 7	2	4	0	0	24	8
Run 8	1	7	0	0	35	7
Run 9	2	10	0	0	35	5
Run 10	2	10	0	0	28	13
Run 11	3	1	0	0	14	12
Htt17_α_slipids	1 ± 1	11 ± 1	0 ± 0	0 ± 0	31 ± 3	19 ± 2
Run 1	1	10	0	0	34	17
Run 2	2	12	0	0	27	22
Htt17_α_hrex	1 ± 1	5 ± 3	0 ± 0	0 ± 0	42 ± 9	4 ± 4
Htt17_α_hrex_2	2 ± 3	5 ± 3	0 ± 0	0 ± 0	53 ± 22	2 ± 2
Htt17_α_hrex_3	1 ± 1	4 ± 2	0 ± 0	0 ± 0	29 ± 16	5 ± 3

Probability of occurrence of intramolecular salt-bridges in Htt17. These quantities are computed over the 250–1000 ns time interval for the MD simulations. The first line for each simulation type combines the averages of all MD runs. For the HREX simulations, the average and standard deviation are computed at the unscaled replica on the 50–250 ns time interval using 20-ns time windows.

Table V.V – Orientation of Htt17 on the bilayer.

	^{15}N-Leu^{7} ppm	$^{2}H_3$-Ala^{10} kHz	^{15}N-Phe^{11} ppm	^{15}N-Phe^{17} ppm	Tilt °	Pitch °
Htt17_α_slipids	71 ± 1	17 ± 1	88 ± 1	86 ± 2	88 ± 5	83 ± 5
Htt17_α_pope	68 ± 3	6 ± 6	86 ± 2	98 ± 1	92 ± 5	93 ± 5

The orientation of the membrane anchor is directly related to the ^{15}N chemical shifts of Leu–7, Phe–11 and Phe–17 as well as the $^{2}H_3$ quadrupolar splitting of Ala–10 [5, 36, 408]. The Htt17_α_pope simulation refers to previously published results on the interactions between Htt17 and a POPE bilayer using the AMBER99sb*-ILDN/SLIPIDS forcefield [111]. The value and the error bar for each residue are respectively the average and the standard deviation over all the averages on the 250–1000 ns (POPC) and 250–500 ns (POPE) time intervals obtained from two independent MD simulations.

Table V.VI – Comparison between POPC and POPE in terms of the average number of hydrogen bonds and salt-bridges between Htt17 and the phospholipids of the membrane.

Interaction type	Residue	Htt17_α_slipids no. (%)	Htt17_α_pope no. (%)
Hydrogen bonds	Glu–5	–	1.1 ± 0.1 (33 ± 3)
	Glu–12	–	1.1 ± 0.1 (28 ± 11)
	Thr–3	1.0 ± 0.0 (37 ± 21)	1.0 ± 0.0 (81 ± 9)
	Lys–6	1.3 ± 0.0 (63 ± 4)	1.3 ± 0.1 (49 ± 8)
	Lys–9	1.0 ± 0.0 (9 ± 1)	1.0 ± 0.0 (7 ± 1)
	Ser–13	1.0 ± 0.0 (47 ± 10)	1.0 ± 0.0 (36 ± 12)
	Lys–15	1.3 ± 0.0 (63 ± 1)	1.2 ± 0.1 (58 ± 1)
	Ser–16	1.0 ± 0.0 (10 ± 1)	1.0 ± 0.0 (8 ± 3)
Salt-bridges	Glu–5	1.0 ± 0.0 (3 ± 1)	1.0 ± 0.0 (34 ± 3)
	Glu–12	1.0 ± 0.0 (2 ± 1)	1.1 ± 0.0 (29 ± 10)
	Lys–6	1.2 ± 0.0 (51 ± 3)	1.2 ± 0.0 (39 ± 8)
	Lys–9	1.0 ± 0.0 (10 ± 1)	1.0 ± 0.0 (6 ± 1)
	Lys–15	1.1 ± 0.0 (46 ± 1)	1.1 ± 0.0 (46 ± 3)

Comparison between POPC and POPE in terms of the average number of hydrogen bonds and salt-bridges between Htt17 and the phospholipids of the membrane. The numbers in parenthesis represent the probability of occurrence in percentage. The Htt17_α_pope simulation refers to previously published results on a POPE bilayer using the AMBER99sb*-ILDN/SLIPIDS forcefield [111]. The value and the error bar for each residue are respectively the average and the standard deviation over all the averages on the 250–1000 ns (POPC) and 250–500 ns (POPE) time intervals obtained from two independent MD simulations.

Annexe VI

Le logiciel de simulations opep_sim pour la famille de champs de force OPEP

A molecular dynamics program for the family of forcefields OPEP

Sébastien Côté, Vincent Binette, Laurent Jutras-Dubé
and Normand Mousseau

Département de Physique and Groupe d'étude des protéines membranaires (GÉPROM), Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec), Canada

Manuscrit en préparation. L'inclusion dans le code de nouveaux algorithmes tels que l'échange de répliques et la parallélisation du calcul des forces ainsi que la caractérisation de la rapidité du code seront effectués dans le groupe du professeur Normand Mousseau. Cette section a été écrite dans l'optique d'une publication future.

Abstract

The investigation of protein structure, folding and aggregation requires a forcefield to model the physical interactions in proteins. Coarse-grained forcefields are a necessary simplification to access longer timescales, large systems and faster sampling than what is typically achievable using all-atom explicit solvent forcefields. The optimized potential for efficient structure prediction (OPEP), for instance, is a coarse-grained forcefield that has been applied on numerous problematics of interest ranging from structure prediction, protein folding to amyloid protein aggregation. The core philosophy of OPEP is based upon the following : (i) it is solvent- and charge-free, (ii) it has specific non-bonded parameters for each side-chain/side-chain pair to take into account their hydrophobic and charged interactions and (iii) it has explicit and cooperative H-bonds. Since this latter interaction type is three- and four-body, it is difficult to include OPEP in the open source and highly parallelized softwares currently available. Here, we introduce the new molecular dynamics software opep_sim that has been developed and optimized for the OPEP family of forcefields – OPEPv3, OPEPv4, sOPEP and aaOPEP – and their future developments. We present the structure of opep_sim and of its topology generator as well as the implemented algorithms and interaction functions.

Material and Methods

In this study, we present a molecular dynamics software called opep_sim that is designed and optimized for the family of forcefields OPEP. Below, we describe the file system of the topology generator and of opep_sim. We also present the algorithms and the interaction functions implemented in opep_sim. Finally, we describe the validation protocol that has been applied to confirm that the family of forcefields OPEP have been correctly implemented in opep_sim.

The topology generator. One of the main pillars of opep_sim is the flexibility of its topology generator for the future developments of the family of forcefields OPEP. The topology generator converts a regular PDB file to a topology file (.top) and a new PDB file that are compatible with opep_sim. The topology file contains all the necessary information – all bonded and non-bonded parameters and participating atoms – to describe the molecular system to be simulated by opep_sim. The topology generator uses four parameter files for each forcefield version. They specify (i) the molecular structure of each amino acid (or any compounds for that matter) in the file with the *rtp* extension, (ii) the bonded and non-bonded parameters in the file with the *ffp* extension, (iii) the exclusion rules to remove or modify a specific non-bonded interaction in the file with the *xcl* extension, and (iv) the rules to change the name of atoms and to convert groups of atoms to pseudoatoms (coarse-graining) in the file with the *mpg* extension. The currently available line types for each file (rtp, ffp, xcl and mpg) are summarized in Table VI.II. The format for each line type is presented in Table VI.III.

The topology generator works as follow. First, the PDB file containing the molecular system as well as all topology source files (rtp, ffp, xcl and mpg) are read. Second, the information in the rtp file is used to confirm that there isn't any atom missing in the PDB file. It is also used to set the covalent bonds and improper torsion angles. The information in the mpg file is used when there are missing atoms to see if they can be found under other names or if they need to be built (coarse-graining). Third, all torsion angles, which corresponds to all quartet of atoms linked by three consecutive covalent bonds, are found from the molecular structure. Fourth, the bonded parameters from the ffp file are associated to the bonds, bond angles, torsion angles and improper torsion angles of the molecular structure. Fifth, the non-bonded parameters from the ffp file – van der Waals, hydro, hydrogen bond and electrostatic – are computed for each atom type pair. Sixth, the information in the xcl file is used to completely ignore some interactions and to replace some of them with another interaction. Finally, a topology file (top) is generated.

The topology file contains in sequential order (i) the number, the name, the residue name, the chain number, the residue number, the mass, the charge, the van der Waals index type, the hydro index type, the hydrogen bond index type and the electrostatic index type for each atom, (ii) all bonds in the molecular system and their type number, (iii) all bond types with their function type and parameters, (iv) all bond angles in the molecular system and their type number, (v) all bond angle types with their function type and parameters, (vi) all torsion angles in the molecular system and their type number, (vii) all torsion angle types with their function type and parameters, (viii) all improper torsion angles in the molecular system and their type number, (ix) all improper torsion angle types with their function type and parameters, (x) the parameters for all van der Waals atom type pairs, (xi) the parameters for all hydro atom type pairs, (xii) the parameters for all H-bond atom type pairs, (xiii) the parameters for all cooperative H-bond types, (xiv) the parameters for all electrostatic atom type pairs, (xv) all special interactions as specified by the xcl file and their type number, (xvi) all special interaction types with their function type and parameters, (xvii) the list of atoms that are excluded from van der Waals interactions, (xviii) the list of atoms that are excluded from electrostatic interactions, and (xix) the list of atoms that are excluded from H-bond interactions.

This topology generator completely uncouple the parameters handling from the simulation program, which is not the case in the previous OPEP program that has a lot of hardcoded parameters. The new file system allows a greater flexibility : the implementation of a new potential requires to only create a new set of rtp, ffp, xcl and mpg files, given that the function types of the potential are already implemented in opep_sim. The currently available function types for each line type are summarized in Table VI.IV. Basically, the functions in OPEPv3, OPEPv4, sOPEP and aaOPEP potentials are already implemented.

The file system of opep_sim. The opep_sim program requires three input files : (i) the PDB file and the topology file created by the topology generator and (ii) the input simulation parameters file with the *isp* extension. The PDB file is used to set the initial position of each atom in the system, while the topology file is used to set the bonded and non-bonded parameters, interactions and exclusions. Finally, the isp file specifies the simulation parameters to perform an energy minimization, a thermalization or a molecular dynamics simulation. A restraint file (*rst* extension) can be read by opep_sim to set up, if needed, the restraints used during the simulation. The restraints can be the position of an atom, the distance between two atoms, the angle between

three atoms and the dihedral angle between four atoms.

The algorithms in opep_sim. The algorithms implemented in opep_sim for energy minimizations and molecular dynamics simulations are presented in Table VI.I. Energy minimization can be performed using three different algorithms : two that use a line descent protocol (SD and CG) and one that uses a damped–stimulated molecular dynamics simulation (FIRE). This latter algorithm improves the convergence in meaningful minima by not converging to shallow minima. Molecular dynamics simulations can be performed in the NVE ensemble or in the NVT ensemble using the Berendsen thermostat. Constraints can be applied to the bond length during a simulation to increase the time step of the velocity Verlet integrator using either RATTLE or LINCS.

Table VI.I – Algorithms implemented in opep_sim

Use	Algorithm name	Reference
Energy minimization	Steepest Descent (SD)	–
	Conjugate Gradient (CG)	–
	Fast Inertial Relaxation Engine (FIRE)	[56]
Integrator for molecular dynamics	Velocity Verlet	[452]
Thermostats	Berendsen	[40]
Constraints	RATTLE	[9]
	LINCS	[184]

The implementation of the family of forcefields OPEP. We refer to previous publications for the functional form of OPEPv3 [311], OPEPv4 [89], sOPEP [310] and aaOPEP [107]. To confirm that these models are correctly implemented in opep_sim, we first verified that the negative gradient of the energy indeed corresponds to the force for each interaction in various proteins and structures. We also compared the potential energies computed using opep_sim to those computed by other simulation programs. We consider that our implementation is valid when all the main energy terms have a relative error of less than 0.001 percent in these comparisons.

The main energy terms for OPEPv3 are the bond, bond angle, torsion angle, phi–psi restraint, regular Lennard-Jones, side-chain/side-chain, C_α–C_α, H-bond and cooperative H-bond interactions [311]. We compared the values given by opep_sim to those given by the previous software developed specifically for OPEPv3. To get a good agreement, we had to correct a few inconsistencies of this software. First, for the C_α of two non-neighboring glycines, the only energy computed is the side-chain/side-chain interaction. Before, the regular Lennard-Jones, the

side-chain/side-chain and, if separated by more than 4 residues, the C_α–C_α interactions were computed. We keep only the side-chain/side-chain interaction energy as it dominates the two others. Second, for the C_α of two neighboring glycines, the only energy computed is the regular Lennard-Jones interaction. Before, both the regular Lennard-Jones and the side-chain/side-chain interactions were computed. We keep only the former because the latter is repulsive for neighboring side-chains while the two C_α are linked by a torsion angle. Third, for all other C_α that are spaced by more than 4 residues, the only energy computed is the C_α–C_α interaction. Before, both the regular Lennard-Jones and the C_α–C_α interactions were computed. We keep only the latter because the former is negligible with respect to it. Fourth, for the H-bond cooperativity, only the strongest H-bond for a given hydrogen of a donor can cooperate. Before, the H-bonds chosen to cooperate depended on the order used to compute the H-bonds as well as an energy threshold that was not very restrictive.

The only change introduced in OPEPv4 with respect to OPEPv3 are the function for the side-chain/side-chain interactions as well as the presence of eleven new i/i+3 and i/i+4 side-chain/side-chain specific interactions [89]. We compared the energies computed by opep_sim to those given by the simulation software specifically developed for OPEPv4. To get a good agreement, we have corrected the same inconsistencies as for OPEPv3. Moreover, we had to modify the $G(r_{ij}^0)$ function presented in the publication describing the OPEPv4 potential [89] because it is different from the implemented function in the software. In opep_sim, we use the function implemented in the software : $G(\sigma) = -0.7 \exp\left[2\left(\frac{\sigma-0.5}{5.0}\right)\right](\sigma - 0.5)$ where σ is in Angstroms.

The main energy terms of sOPEP are the regular Lennard-Jones, side-chain/side-chain, C_α/C_α, H-bond and cooperative H-bond interactions [310]. We compared the values given by opep_sim to those given by the PEPFOLD program. The bonded terms for sOPEP in opep_sim are taken from OPEPv3. This is necessary to perform energy minimizations and molecular dynamics simulations. The same corrections for OPEPv3 are made to sOPEP.

The potential aaOPEP is based upon both AMBER99sb*-ILDN [382] and OPEP [107]. The bond, bond angle, torsion angle and regular Lennard-Jones interactions come from AMBER99sb*-ILDN, so we compare their energy computed by opep_sim to those computed by GROMACS 4.5.4 [387]. The remaining interactions – H-bond and cooperative H-bond energies – come from OPEP, so their implementation has already been confirmed when we tested OPEPv3. and

OPEPv4.

Table VI.II – Line types in the files used by the topology generator.

File extension	Line type	Use
rtp	MODEL	Starts the molecular structure of an amino acid
	ATOM	Declares the name of an atom that is in the amino acid as well as its atom type, charge and ability to form a H-bond
	BOND	Declares a covalent bond between two atoms
	ITOR	Declares an improper torsion angle to maintain a particular geometry between four atoms
	END	Ends the molecular structure of an amino acid
ffp	BOND	Declares the function type and parameters for a bond type
	BANG	Declares the function type and parameters for a bond angle type
	TORA	Declares the function type and parameters for a torsion angle type
	ITOR	Declares the function type and parameters for an improper torsion angle type
	LJINT	Declares the mass, function type and van der Waals parameters for an atom type
	LJSPC	Declares the function type and van der Waals parameters for an atom pair by overriding the value calculated from LJINT
	HYDRO	Declares a scaling factor to the van der Waals interactions for an amino acid pair type
	HBINT	Declares an atom pair that can form an H-bond as well as the associated function type and parameters
	ESINT	Declares a pair of atoms that can form a salt-bridge as well as the associated function type and parameters
xcl	LJEXCL	Declares an exclusion rule to remove or modify a specific van der Waals interaction between an atom pair
	HBEXCL	Declares an exclusion rule to remove or modify a specific H-bond interaction between an atom pair
	ESEXCL	Declares an exclusion rule to remove or modify a specific electrostatic interaction between an atom pair
mpg	(none)	Each line specifies a rule to change an atom name or to convert a group of atoms to a single atom (coarse-graining)

Table VI.III – Format for each line type in the files of the topology generator.

File extension	Line format
rtp	MODEL aaname$^{(a)}$ number_atom number_bond number_itor
	ATOM atomname atomtype partial_charge *obsolete_opt* hbond_type$^{(b)}$
	BOND atomname1 atomname2 residue_ndx1$^{(c)}$ residue_ndx2$^{(c)}$
	ITOR atomname1 atomname2 atomname3 atomname4 residue_ndx1$^{(c)}$ residue_ndx2$^{(c)}$ residue_ndx3$^{(c)}$ residue_ndx4$^{(c)}$
	END
ffp	BOND atomtype1 atomtype2 function_type parameters$^{(d)}$
	BANG atomtype1 atomtype2 atomtype3 function_type parameters$^{(d)}$
	TORA aaname1$^{(a)}$ aaname2$^{(a)}$ aaname3$^{(a)}$ aaname4$^{(a)}$ atomtype1 atomtype2 atomtype3 atomtype4 function_type parameters$^{(d)}$
	ITOR *Idem*
	LJINT atomtype mass (amu) function_type parameters$^{(d)}$
	LJSPC atomtype1 atomtype2 function_type parameters$^{(d)}$
	HYDRO aaname1$^{(a)}$ number_atoms1$^{(e)}$ all_atomnames1$^{(f)}$ aaname2$^{(a)}$ number_atoms2$^{(e)}$ all_atomnames2$^{(f)}$ scale$^{(g)}$ *obsolete_opt*
	HBINT aaname1$^{(a)}$ atomname1$^{(h)}$ aaname2$^{(a)}$ atomname2$^{(h)}$ function_type parameters$^{(d)}$
	ESINT *Idem*
xcl	LJEXCL aaname1$^{(a)}$ aaname2$^{(a)}$ delta_resndx$^{(i)}$ atomname1 atomname2 delta_atomndx$^{(j)}$ function_type$^{(k)}$ parameters$^{(d)}$
	HBEXCL *Idem*
	ESEXCL *Idem*
mpg	new_atomname instruction_number$^{(l)}$ atomname(s)

(a) Amino acid name that corresponds to a N- or C-terminal amino acid when the name begins by *n* or *c*, respectively. (b) H-bond type : 1 means a hydrogen linked to a donor, -1 means an acceptor and 0 means neither. (c) Residue index : -1 means atom from previous residue, 0 means atom from current residue, 1 means atom from next residue and so forth. (d) The format and the number of parameters depend on the function_type, see Table VI.IV. (e) Number of atoms in the corresponding amino acid name for which their van der Waals interactions are scaled by the HYDRO parameters. (f) The name of the atoms to be scaled for the associated amino acid name. (g) The scale multiples the van der Waals interaction of all atomname1/atomname2 specified pairs. (h) The atom name is associated to its corresponding amino acid name. (i) The number of residues between aaname1 and aaname2 in the same polypeptide sequence where -1 means any number. (j) The number of covalent bonds between atomname1 and atomname2 in the same polypeptide sequence where -1 means any number. (k) Here, the function types available are the same as the corresponding interactions (LJ, HB or ES) ; except when it is zero meaning that this interaction is ignored or when it is negative meaning that the declared parameters are scaled. (l) The instruction number can be zero to just replace a single atom name by the new atom name or one to replace the specified atoms by a pseudoatom at their center-of-mass when coarse-graining.

Table VI.IV – Function types for each line type in the files of the topology generator.

File extension	Line type	Function type	Function	Parameter order
ffp	BOND	1	$k_r\left(r-r_{\mathrm{eq}}\right)^2$	k_r (kcal/mol/Å^2), r_{eq} (Å)
	BANG	1	$k_\theta\left(\theta-\theta_{\mathrm{eq}}\right)^2$	k_θ (kcal/mol/deg^2), θ (deg)
	TORA	1	$k_n\left(1-\cos\left(n\phi-\delta_n\right)\right)$	δ_n (deg), k_n (kcal/mol), n
		$3^{(a)}$	$k_n\left(1-\cos\left(n\phi-\delta_n\right)\right)+k_\phi\left(\phi-\phi_0\right)^2$	δ_n (deg), k_n (kcal/mol), n, k_ϕ, ϕ_{lower} (deg), ϕ_{upper} (deg)
	ITORS	1	$k_n\left(1-\cos\left(n\phi-\delta_n\right)\right)$	δ_n (deg), k_n (kcal/mol), n
	LJINT	1	$\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12}-2\left(\frac{\sigma}{r}\right)^6\right]$	σ (Å), ε (kcal/mol)
		4	$\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12}-2\left(\frac{\sigma}{r}\right)^6\right]\mathrm{H}(\varepsilon)-\varepsilon\left(\frac{\sigma}{r}\right)^6\mathrm{H}(-\varepsilon)$	σ (Å), ε (kcal/mol)
		$6^{(b)}$	$U_{\mathrm{DB}}\left(\varepsilon,\varepsilon_{db},\varepsilon_{ssm},r_{cm}\right)$	ε (kcal/mol), ε_{db} (kcal/mol), ε_{ssm} (kcal/mol), r_{cm} (Å)
		$7^{(c)}$	$\varepsilon\left[\left(\frac{\sigma-p}{r-p}\right)^{12}-2\left(\frac{\sigma-p}{r-p}\right)^6\right]\mathrm{H}(\varepsilon)-$ $\varepsilon\left(\frac{2R_0-\sigma}{r}\right)^6\mathrm{H}(-\varepsilon)\quad$ where $\quad p=\frac{\sigma-\sqrt[6]{2}R_0}{1-\sqrt[6]{2}}$	R_0 (Å), σ (Å), ε (kcal/mol)
		$8^{(d)}$	$\varepsilon\left[\left(\frac{G(\sigma)}{r}\right)^6\mathrm{e}^{-2r}+0.6563701\cdot\right.$ $\left.(\tanh\left[2\left(r-\sigma-0.5\right)\right]-1)\right]\mathrm{H}\left(\varepsilon\right)-\varepsilon\left(\frac{\sigma}{r}\right)^8\mathrm{H}\left(-\varepsilon\right)$	σ (Å), ε (kcal/mol)
	LJSPC		Same as LJINT	
	HBINT	100	$\varepsilon\left[5\left(\frac{\sigma}{r}\right)^{12}-6\left(\frac{\sigma}{r}\right)^{10}\right]\cos^2\left(\alpha\right)\mathrm{H}\left(\alpha-90^\circ\right)$	σ (Å), ε (kcal/mol), hbcoop_fcttype
	ESINT	50	kq_iq_j/r	k (kcal·Å/mol/e^2)
xcl	LJEXCL		Same as LJINT	
	HBEXCL		Same as HBINT	
	ESEXCL		Same as ESINT	

(a) Regular torsion angle augmented by the phi–psi potential (OPEP) to restrain the dihedral angles ϕ and ψ to the relevant regions of the Ramachandran plot : $\phi_0=\phi_{\mathrm{upper}}$ when $\phi>\phi_{\mathrm{upper}}$ and $\phi_0=\phi_{\mathrm{lower}}$ when $\phi<\phi_{\mathrm{lower}}$ [311]. (b) Van der Waals interaction with a desolvation barrier (U_{DB}) with $r_{ssm}=r_{cm}+3$, $r_{db}=(r_{ssm}+r_{cm})/2$, $k=6$, $m=3$ and $n=2$ as previously presented and justified by Hue Sun Chan and co-workers [84, 95, 107, 226]. (c) The side-chain/side-chain interaction in sOPEP [310]. (d) The side-chain/side-chain interaction in OPEPv4 [89].

Perspective

Le but de opep_sim est d'offrir une plateforme de développement pour la famille de champs de force OPEP. Précédemment, une nouvelle version du logiciel de simulation OPEP devait être produite à chaque nouvelle version de OPEP parce que (i) certains paramètres sont directement inclus dans le code et (ii) le code ne permettait pas facilement l'utilisation de plusieurs fonctions d'interaction différentes. De plus, la liste de toutes les interactions liées et non-liées était gardée explicitement en mémoire et la réutilisation des données en cache était marginale. S'ajoute à cela le fait que l'ancien générateur de topologie n'est pas très flexible, car il consistait en un ensemble de scripts et de bouts de codes jumelés au générateur de topologie de AMBER91.

Maintenant, nous avons un générateur de topologie qui découple entièrement du logiciel de simulation la préparation des paramètres d'interaction. De plus, le nouveau générateur de topologie est très flexible – compatible avec OPEPv3, OPEPv4, sOPEP et aaOPEP – et il permet très facilement l'ajout de nouveaux modèles. Le logiciel opep_sim trouve à la volée les paires d'atomes participant aux interactions non-liées grâce à la séparation de la boîte périodique en sous-boîtes qui est rendue possible grâce à l'emploi de distances de troncation. Ceci permet une meilleure utilisation de la mémoire cache. De plus, il est compatible avec une variété de fonctions d'interaction et il permet l'application de restreintes variées sur les atomes du système. Finalement, l'ajout de nouvelles fonctions d'interaction ou de nouveaux algorithmes dans opep_sim est simple et ne requiert que des changements très minimes au code principal.

Les limitations actuelles de opep_sim sont : (i) l'absence d'un algorithme d'échantillonnage avancé tel que la dynamique moléculaire d'échanges de répliques [449] et *simulated tempering* [346], (ii) l'absence d'un thermostat générant correctement l'ensemble canonique tel que le thermostat Bussi-Donadio-Parrinello [75] et l'absence d'une dynamique de Langevin [439], ainsi que (iii) l'absence de la parallélisation du calcul des forces en utilisant les listes de Verlet [481]. Le développement de opep_sim se poursuivra dans le groupe du professeur Normand Mousseau.

Contributions

S.C. designed and coded the current version of the software ; S.C. designed and coded the topology generator ; S.C. implemented OPEPv3, sOPEP and aaOPEP ; L.J.D. implemented OPEPv4 ;

S.C., V.B. and L.J.D. tested the software ; S.C. and N.M. wrote the manuscript.

Acknowledgments

This work was funded in part by the Canada Research Chairs program, the Fonds de recherche du Québec – Nature et Technologies (FRQNT), the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Fonds de recherche du Québec – Santé (FQRS). The authors thank Jonathan Ferland of Calcul Québec for designing the data structure of the code.